Preliminary Comments from Members of the Radiation Advisory Committee Augmented for the Review of the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM), Revision 2, Public Comment Draft

Comments Received as of January 6, 2021

Table of Contents

Dr. Sally Amundson

General Comments:

- I'm confused by the last decision point on the flowchart in Fig 2.5 (and Fig. 3.1). It seems to indicate that if the site may contain radioactive material in excess of background, it should be released, but if there is no possibility of elevated levels it moves forward to the investigation process in fig 2.6. This decision node has been reworded from Rev 1, apparently reversing the meaning of the question, yet the yes and no arrows still point in the same direction as in Rev1. Should yes and no be reversed here or am I missing something?
- As figure 4.2 is both downstream and upstream of figure 4.1 is there a way it can be brought closer to fig 4.1? Having to shuffle 30 pages ahead to follow the logic flow in 4.1 seems disruptive. It's unclear why fig 4.2 needs to be where it is. The bubbles could also usefully refer forward to the appropriate measurement sections as was done in Fig. 4.1 and other flowcharts.

Charge 3.1

- In section 4.12.3.1 where LBGR is first used it seems confusing that the examples must rely on a table in the next chapter. Might it be possible to integrate most of this information into chapter 5, where it is also covered, with a higher-level summary of sample size determination being presented in chapter 4, and reference (as is already given) to the appropriate sections in chapter 5? I'm uncertain of the demarcation between chapters 4 & 5. It seems that the examples in chapter 4 might be integrated into chapter 5, where there seems to be some overlap anyway?
- The point that using ½ DCGLw as the LBGR is not an ideal approach is made quite clearly on p.5-28 ll.15-16. It also raises the expectation that if this is done, it will only be used as a first pass estimate to be refined later. I think the example boxes (4 & 6) are also really useful in this regard, as they illustrate how the LBGR is set using information from the specific site.
- Generally the introduction of LBGR p.2-12 ll.14-17 is a great improvement over the Rev1 definition that started out saying it was defined as 50% DCGLw and then adjusted. Looking at Rev1, I couldn't find a clear explanation of how to determine (or adjust) the LBGR, it just sort of springs up in use, so perhaps not surprising that the definition of 50% DCGLw was broadly adopted.
- In example 5 (p.8-23), perhaps adding a phrase such as "Based on the site survey measurements," before "the LBGR was selected to be…." To re-emphasize that this is site-specific and should not be expected to be an arbitrary or "cook-book" number. The same comment for p.8-25, example 6 and p.8-30, example 7

Charge 3.2

- Yes, I find the description in Rev2 easier to follow, in particular because in Rev1 the Area Factor wasn't described in that section, but referred back to an earlier section. Reading the description in Rev2 I had the impression that the EMC release criteria were an integral part of the assessment plan developed for the individual site, rather than something generic that could be obtained from the literature.
- Also, the movement of the more detailed discussion of Area Factor use to Appendix O is helpful in this regard.

Charge 3.4

- I like the addition of the flowcharts (figs 5-1 to 5.3). They nicely summarize how all the different survey types work together for the different class sites.
- Increased use of bullets rather than a wall of text is also helpful to the reader (everywhere, but especially in Chapter 5)
- Selecting the Scenario and the description of how to determine LBGR fit well into the flow and seem to make logical sense here.
- On p.5-31 l.7 it would be useful to add text like that used in describing Table 5-3 to the description of table 5-2 "These values were calculated using Equation O-1 in Appendix O and increased by 20% to account for missing or unusable data and uncertainty in the calculated value of N."
- I do like offloading much of the derivations associated with determining the number of samples to the appendix. Not only does it make the main text more readable, but also it makes the examples of actually performing the calculations using the tables very clear. In practice, people would presumably use the tables, rather than deriving the values each time. It is important to make this information available, but use of the appendix both streamlines reading of the main text to improve comprehension of the method, and makes it a more practical guide.

Dr. Roland Benke

ASSIGNED CHARGE QUESTIONS

1) Are the revisions to MARSSIM concepts and methodologies technically appropriate, useful and clear, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

1.1 Please identify whether the inclusion and proposed implementation of scan-only surveys (Section 5.3.6.1 and Section 8.5) is appropriate, adequate and clear, especially the discussion on sampling for scan-only measurement method validation or verification.

Section 8.5 for scan-only surveys includes Equation 8-3 for an upper confidence limit (UCL). The UCL is described as the parameter of interest for demonstrating compliance (Section 2.5.1.2). Overlapping fields of view are identified for a series of *in situ* measurements in Section 8.2.3 and Table 8.1. The proposed approach serves as a good starting point but requires more discussion on instrument response and calibration factors to be adequate. Overlapping fields of view should be addressed for scan-only surveys with gamma-ray detectors.

Some scan-only systems allow for numerous measurements of surveyed areas with overlapping fields of view. For penetrating gamma rays, the inferred concentration from a given instrument response can vary considerably over the detector's field of view. Systems with multiple sensors also exhibit nonuniform spatial response profiles. Examples presently in MARSSIM assume the detection system is positioned directly over the center of contamination and does not respond to nearby contamination in adjacent areas. However, this assumption does not hold in the field when scanning large areas with gamma-ray detectors because nearby areas of elevated contamination contribute to detector measurements. Quantifying the spatial response of the system will avoid problems with statistical inferences from gamma-ray scanning measurements with manual, mobile, and aerial scan-only survey equipment featuring uncollimated detectors. Although the field of view can be restricted with collimation shielding, uncollimated gamma-ray equipment is more popular for large area surveys. In contrast, scanning instruments for beta and alpha particles generally respond to source material immediately below the detector, and thus, treatment for overlapping fields of view and nonuniform spatial response is unnecessary for scan-only surveys with those instruments.

When the Scan MDC is satisfactory for a given objective, scanning measurements can be very effective. The expectation for complete 100-percent coverage with scan-only surveys, per Table 5.5 and Section 5.3.6.1 for Class 1 areas, should not be interpreted to rule out scanning surveys performed on smaller portions of a survey unit when coupled with other surveys for remaining areas. For scanning measurements with mobile detection systems, the most effective approach is to perform on-site confirmation of instrument's detection capability prior to survey measurements. Section 6.3.2 presents statistical formulations tied to decision making for audible scanning measurement without data logging for a two-staged measurement technique, performing a scanning measurement to locate suspected areas of elevated concentrations and pausing to acquire a short-duration static measurements. Maps of radiation data generated by data-logging scanning systems that track real-time position represent the greatest opportunity for MARSSIM guidance advancement. These rich data sets capture a continuous stream of measurements primarily while the detection system is in motion. Designed for systems without data logging, the two-staged technique with frequent static measurements doesn't provide the same benefits for modern systems with data logging of detector location and response. Establishing a scan-only protocol with (1) preselected locations for static measurements according to a grid, (2) continuous scanning measurements at a constant speed, and (3) follow-up static measurements at locations suspected to have the highest concentrations would maximize the efficiency and utility of collected survey data. Separate MDC calculations for scanning and static measurements by the same system would remain.

Once quality objectives have been satisfied, the need for sampling and laboratory analysis to confirm scan-only surveys is chiefly related to calibration of the scanning equipment. When the site-specific environment is considered by calibrating the scanning equipment in the field with the contaminant of interest, sampling and laboratory analysis are not required for verification. When the scanning equipment is calibrated in the laboratory or with computational methods, sampling and laboratory analysis are recommended for verification. Static *in situ* measurements can be used in place of sampling, described in Section 5.3.6.2, provided that the static *in situ* measurement can quantify the average amount of radioactive material in the survey unit.

Given that $DCGL_{EMC}$ cannot be less than $DCGL_w$, an inconsistency appears with recommendations for the Scan MDC. Different MDC recommendations for scan-only surveys may be responsible. Refer to the excerpts below.

- On page 4-18: "For scanning, minimum detectable concentration (MDC) considerations are a little more nuanced than for discrete sampling. For example, when scanning for areas with potentially elevated concentrations of residual radioactive material, the scan MDC should be below the DCGL—preferably at a fraction (approximately 50 percent) of the DCGL. In a Class 1 survey unit, the scan MDC should be less than the $DCGL_{EMC}$."
- Equation 5-5 indicates the required Scan MDC equals $DCGL_{EMC}$.
- On page 5-43 for scan-only surveys, MARSSIM currently recommends that the Scan MDC should be less than 50 percent of the $DCGL_w$.
    - On page 5-52: "… a scan-only approach should only be used for circumstances where the scan MDC (for the scan system) is less than 50 percent of the $DCGL_w$ and other MQOs, such as requirements for measurement method uncertainty, can be met."
    - On page 5-53 for Class 1 areas: "As discussed in **Section 5.3.6.1,** scanning techniques can be used in lieu of discrete samples or direct measurements when the scan MDC is less than 50 percent of the $DCGL_w$, and the scan coverage is 100 percent."
    - On page 5-53 for Class 2 areas: "As discussed in **Section 5.3.6.1,** scanning techniques for Class 2 survey units might be used in lieu of discrete samples or direct measurements when the scan MDC is less than 50 percent of the $DCGL_w$ and the scan coverage is between 10 and 100 percent."

Compared to guidance for other survey techniques, MDC recommendations seem more stringent (by a factor of 2) for scan-only surveys, which is counterintuitive given their larger coverage and greater confidence against missing contaminated regions. If more stringent levels are indeed necessary for scan-only surveys, a technical basis to support that position should be presented.

1.4 Is the discussion of survey requirements for areas of elevated activity technically accurate, appropriate and clear? In particular, please comment on the decision to maintain the use of the unity rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4). Are there suggested alternatives to the use of the unity rule?

Charge Question 1.4 asks for suggested alternatives to the unity rule for elevated radioactive measurements. Area-weighting of elevated radioactive measurements with other survey results is recommended because it facilitates combination of these attributes with a representation that maintains proportionality to dose and risk without implying exposures are exclusively centered over elevated radioactive measurements. Double counting of areas should be avoided, especially during the incorporation of data from additional or follow-up surveys.

Equation 8-4 is not advised. As suggested in the text below the equation, adding additional terms for multiple areas of elevated radioactive material is an “impossible situation and represents a very cautious exposure scenario.” Such an unrealistic, overly conservative tendency should be removed from the guidance. Fortunately, good alternatives already suggested in Section 8.6.2 for addressing multiple areas of elevated radioactive material can be presented without the equation. Coordination with the regulator is endorsed, but regulatory uncertainties should not prevent the process from moving forward. Section 4.5.2 indicates that the implementer can justify to the regulator when no $DCGL_{EMC}$ requirement is necessary. Areas with elevated radioactive measurements including aspects on deriving $DCGL_{EMC}$, which are beyond the scope of MARSSIM, have and can result in confusion and misinterpretation.

To determine when a $DCGL_{EMC}$ evaluation is necessary, an area-weighting approach—as suggested above—or other alternative with an unbiased dose and risk perspective is recommended. Developing and vetting such an approach in multi-agency guidance would alleviate past issues with $DCGL_{EMC}$.

1.5 Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles technically sound and appropriate, and is the description accurate? In particular, please comment on the rule-of-thumb for determining when use of MARSSIM may not be appropriate for survey units containing discrete radioactive particles (Section 4.12.8 and Appendix O.5).

Historical motivation for discrete radioactive particle treatment is strongly linked to personnel contamination, localized external doses, and potential intake following deposition on the person, clothing, or personal protective equipment. Considering a MARSSIM emphasis on building surfaces and surface soils, on-site worker activities associated with surveying, decontamination, and release activities provide the best indicator of potential risks. Discrete radioactive particle treatment should become a consideration for the site-specific consideration under MARSSIM only if discrete radioactive particles are actually detected, mitigated, and controlled as part of the operational radiation protection program at the site.

Mentioning inappropriate uses of the MARSSIM EMC process for discrete radioactive particles in Section 4.12.8 is fine. Given that developing guidance on surveys for discrete radioactive particles is not necessary for this revision, ideally MARSSIM could still introduce a rule of thumb when further regulatory and implementer attention on discrete radioactive particles is warranted. Without evidence that discrete radioactive particles represent a credible contribution to dose and risk, ALARA by itself does not constitute a basis for requiring consideration of discrete radioactive particles.

2) Does MARSSIM, Revision 2 provide useful, appropriate and clear examples and descriptions of technical approaches to implementing surveys and the statistics by which they are interpreted?

2.1 Please comment on whether the description of updated measurement methods and instrumentation information (Chapter 6 and Appendix H) is useful, appropriate and clear.

As described on page 6-17, the surveyor efficiency was estimated between 0.5 and 0.75, and MARSSIM recommends a value toward the lower end of this range (i.e., 0.5) for estimating the Scan MDC. Scanning surveys of land areas are addressed within Section 6.3.2, and $MDCR_{surveyor}$ is infused in a majority of the equations and examples. However, it should be clarified that $MDCR_{surveyor}$ is not necessary for data-logging detection systems. Because data logging is included and available in modern detection systems, current guidance pertaining to land area surveys in the reviewed draft is outdated. Many of the fundamental statistical principles still apply, so a minor revision of this section is highly recommended. The statement on scanning techniques in Section 6.6.1.1 reinforces this point, "Scanning equipment coupled with GPS or other locational data is strongly recommended for scan-only surveys." Overall, many important considerations and practical suggestions are communicated in Chapter 6.

Although discussion on unavailable technologies in Appendix H is not immediately useful, it is appropriate and should remain because technological advancements and market forces over time tend to transform some unavailable technologies into widely available technologies. Appendix H provides valuable details on specificity and sensitivity as well as useful summaries of approximated costs and detection limits for field survey and laboratory analysis equipment.

---

3) Is the information in MARSSIM, Revision 2 clear, understandable and presented in a logical sequence? How can the presentation and content of material be modified to improve the understandability of the manual?

3.2. Please comment on whether avoiding the use of the term "area factor" improves understandability of the elevated measurement comparison concept (Section 8.6.1).

Section 8.6 improvements for elevated measurements and comparisons, including the removal of the "area factor" in the description, are endorsed.

3.3 Please comment on the effectiveness of the new organization of Chapter 4 (Considerations for Planning Surveys) to improve the understandability of the Chapter.

The expanded discussion on small areas with elevated activity and discrete radioactive particles, updated flow diagrams, and inclusion of numerous examples are significant improvements toward understandability.

Editorial Comments

The background exposure rate should be corrected in "A typical background exposure rate of 10 mR/h gives a background dose rate of 100 millirem/year (mrem/y)" on page 6-41.

When the true condition is exceeds release criterion, Table D.2 for Scenario B describes a decision error from Accepting H_0 as "Incorrectly Fail to Release Survey Unit." However, this decision error would be incorrectly release survey unit.

When the true condition is meets release criterion, Table D.2 for Scenario B describes a decision error from Rejecting H_0 as "Incorrectly Release Survey Unit." However, this decision error would be incorrectly fail to release survey unit.

On page D-25, line 18: Wording that associates Type II errors with α for Scenario B and β for Scenario A doesn't completely agree with the associations shown in Tables D.1 and D.2.

Page D-25, lines 4-5: … site-specific area factors ~~can also~~**should** be developed.

Dr. Harry Cullings

1.3 Is the proposed implementation of the of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)? Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives? Is the method appropriate and practical for both laboratory and field (including scan) measurements? Please comment on the concerns of stakeholders that calculating measurement uncertainty for field measurements makes the survey process difficult to implement. In addition, please comment on whether recommendations provided by NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).

Re Q1.3, MDC/MDA and MQC/MQA cannot be calculated without implicitly calculating measurement uncertainty, so it seems untenable for a stakeholder not to be able to calculate measurement uncertainty. It is not clear from what is written here what the basis is for the complaints from stakeholders that measurement uncertainty is difficult to calculate for field measurements, particularly if they have to calculate MDC/MDA or MQC/MQA. It seems logical that people performing field measurements are likely to be less well-versed in calculations related to measurement uncertainty than those in counting laboratories, but my question is how they can calculate an MDC/MDA if they cannot calculate the related measurement uncertainty.

Regarding the question of whether the proposed calculation of measurement uncertainty is consistent with the concept of MQO’s, the material given here in Chapter 6 seems to be consistent with that concept, which is to provide a basis for selecting instrumentation and methods to adequately classify sites etc.

I have some concerns with how the material on uncertainty is organized in the manual. Chapter 4, in Sec. 4.8.2, briefly mentions measurement uncertainty as a possible MQO and refers to Chapter 6 in regard to how to calculate it, but does not mention Appendix D at all. Chapter 6, in Sec. 6.4.2, gives a basic calculation of counting uncertainty and in Sec. 6.4.3 gives basic propagation of error formulae, but does not give any other examples, and again does not mention Appendix D. The example shown of how to calculate measurement uncertainty is the simplest and most naïve possible case, and there is only one calculation shown.

In Appendix D, Sec. D4.2.4, there is finally a short list of some things that might be included in measurement uncertainty other than counting Poisson variation. There could be more illustrative examples of how to calculate measurement uncertainty for field measurements, for sources of uncertainty other than Poisson counting uncertainty. The manual in general gives much more detail and illustrative examples in regard to how to calculate MDC/MDA and MQC/MQA than in regard to how to calculate measurement uncertainty.

4.8.2 says the MDC should be less than 50% of the UBGR and refers to Ch 6, but although Ch 6 gives a lot of detail about how to calculate MDCs, it says nothing about how the MDC should compare to the UBGR. The advice that the MDC should be less than 50% of the UBGR is repeated a number of times in the Manual, most often in regard to allowing scanning rather than fixed measurements. There is no mention of "10% to 50% of the UBGR, which is a fairly wide range mentioned in the charge that is otherwise not elaborated upon in the manual. The value that corresponds to a MQC for a 10% fractional standard deviation is somewhere near the middle of this range, but saying the MDC is less than 50% of the UBGR is still not an MDC that is really equivalent to an MQC for a 10% fractional standard deviation.

Since the minimum quantifiable concentration is mentioned in Chapter 6.4, it seems as though some material about how to calculate it might be included in the manual. The material in the MARLAP manual, which is only about 4 pages, seems like a reasonable approach to this. Or, reference could be made to the MARLAP manual, which is readily available online. The MARLAP manual is referenced in regard to calculating measurement uncertainty in Sec. D.1.7.1 of Appendix D, which is good.

It is not clear to me what other material from the noted references (NIST, ANSI/IEEE and MARLAP) could be considered for inclusion in this manual, apart from this material about the MQC.

1.4. Is the discussion of survey requirements for areas of elevated activity technically accurate, appropriate and clear? In particular, please comment on the decision to maintain the use of the unity rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4). Are there suggested alternatives to the use of the unity rule?

Re Q 1.4, the formulation of the unity rule in Sec. 8.4 suggests that the $DCGL_{EMC}$ is a level in addition to, i.e., above, the $DCGL_W$, and that is how it is used in the equation for the unity rule: as an additional separate source above the wide-area source related to the $DCGL_W$. *By this reckoning, the action levels in Table 5.4 should be $DCGL_W + DCGL_{EMC}$, not just $DCGL_{EMC}$.* In any case, the derivation in Sec. 8.4 should be clarified in this regard, i.e., that it is considering 1) the wider area and 2) the portion of the elevated area(s) *above the concentration in the wider area* as the separate sources for the unity rule.

Appendix O in O.4.4 shows tables of example area factors, saying it is strictly for purposes of illustrating the concepts involved, but does not comment on how and why the trends in the tables differ by radionuclide. One assumes this must be due to the ranges of the penetrating radiations emitted, and it would be helpful to at least comment briefly on this and give some relevant additional information related to it.

I think there may be alternatives to the use of the unity rule. For example, the ratios could be calculated once for each site of elevated concentration with the other sites of elevated concentration down-calculated by an application of inverse-square rule based on their distance from the main site under consideration, and the limiting case could be used.

2.2. Please comment on whether the additional optional methodology for the use of Ranked Set Sampling (Appendix E) for hard-to-detect radionuclides is useful, appropriate and clear.

Re Q 2.2, the use of ranked set sampling is amply illustrated but is not explained, as to how it works. The explanations and examples for using it to cull a set of additional samples from a larger set of crude measurements are complicated but useful. However, it would be good to at least include a sentence or two about how the method works in a statistical sense, i.e., by improving the representativeness of the sampling beyond what obtains from purely random sampling. This does not have to get into statistical derivations etc. but it would be intended to give some basic idea of how ranked set sampling works and why it is worthwhile.

There is also very little said about what the correlation coefficient of the crude measurement method that is used for screening must be, in order for the method to work as suggested in the Appendix. It is simply said, rather blandly, in Sec. E.1, that "the screening method selected must have a relative correlation to the concentration of the radionuclide in soil for this procedure to be effective." The crude method must serve to rank the measured locations, but with how much error?

Another point concerns what kind of measurement can be used for the crude screening. The material in Appendix E mentions only the use of in situ measurements of samples with an alpha or beta-detecting probe, whereas the charge material mentions soil particle size as a possible surrogate variable, but that is not mentioned in the manual. What about mentioning some things such as that, and perhaps soil chemistry or other potential surrogates, in Appendix E?

3.1. Please comment on the revised description of how to set the Lower Bound of the Grey Region (LBGR) and its likely effectiveness in encouraging users to rely on site-specific information for doing so (Chapter 4 and Section 5.3).

Re Q 3.1, in terms of statistics the Lower Bound of the Gray Region is technically determined by the accepted Type II error rate beta and the width of the measurement process error distribution. I am not sure how the criterion that it should be taken as a reasonably conservative estimate of the amount remaining in the remediated area was arrived at, and I can find no explanation of it. It would make sense that such a choice would reduce the Type I error as much as possible without increasing the Type II error by getting too far down into the area where the presumed remaining level of contamination is, but this is not explained in any of the materials. There should at least be some verbal description of why a conservatively high estimate of the amount remaining in the remediated area is a good LBGR.

Also, why are there no figures like Fig. 6.1 for the concepts of Type A and Type B decisions? Fig. 6.1 is a very nice figure. There are some figures in one of the appendices, i.e., Figs. D.5 and

D.6, that illustrate some of the basic ideas involved in the Type A and Type B scenarios, but these figures show only the distribution under the null hypothesis and the Type I error; I can find nothing in the manual along the lines of Fig. 6.1 to illustrate the basic quantitative relationships involved in Type A and Type B scenarios.

Editorial comments

For Fig.'s 2.5 and 3.1, the "No" and "Yes" arrows seem to be backwards for both "Does Site Pose Immediate Risk to Human Health and Environment?" and "Does Site Possibly Contain Residual Radioactive Material in Excess of Natural Background or Fallout Levels?"

Page 4-35, line 36: Sec 4.7.3 does not discuss the amount of material to be collected per sample.

Why are some of the DCGL definitions under Abbreviations and others under Symbols, Nomenclature and Notations?

Page numbers are not in line with what is listed in TOC.

Re areas of elevated activity, Sec. 5.3.5 refers to Sec. 4.2.5, which does not exist. Should it refer to Sec. 2.5.3?

Example 8 in 5.3.5.2: shouldn't it say that "The grid area encompassed by a triangular sampling pattern of 10 m is approximately 99.1 m^2," instead of 86.6 m^2?

Appendix D says on page D-6 that MQOs are further developed in Sec. D1.9, but they are in Sec. D1.7.1.

The Table of Contents contains insufficient detail of the sections in the Appendices.

Dr. Lawrence T. Dauer

CHARGE QUESTIONS:

3) Is the information in MARSSIM, Revision 2 clear, understandable and presented in a logical sequence? How can the presentation and content of material be modified to improve the understandability of the manual?

3.1. Please comment on the revised description of how to set the Lower Bound of the Grey Region (LBGR) and its likely effectiveness in encouraging users to rely on site-specific information for doing so (Chapter 4 and Section 5.3).

One of the critical decisions made during site survey design under MARSSIM Scenario A is to set a value for the LBGR. Twenty years of training and review of survey plans have shown that this concept is not well understood by users, and that users tend to implement the standard rule of thumb of setting the LBGR to 50% of the DCGLw. This rule of thumb was provided in MARSSIM, Revision 1, for use only when additional information was not available. A poorly chosen value for the LBGR can affect the power of a survey resulting in unnecessary use of resources or a higher chance of failing a survey unit that meets the release criteria.

In Scenario A, the LBGR should be set equal to a conservative estimate of the average concentration remaining in the survey unit. This information is typically available from historical site information, or a scoping or characterization survey if the survey unit is un-remediated, or the remedial action survey if the site has been remediated. The purpose of the revisions is to describe this concept in plain language, moving away from a statistics terminology description of the concept.

Response:

As noted in the charge question, the setting of the LBGR is a critical decision for the MARSSIM process. The first mention of the LBGR is on page 2-12, Lines 12-19, under ‘Overview of Radiation Survey and Site Investigation Process’. This section mentions that under Scenario A the LBGR is a site-specific variable generally chosen to be conservative (slightly higher) estimate of the concentration of residual radioactive material remaining in the survey unit and adjusted to provide an acceptable value for the relative shift (note that in this section the relative shift has not yet been discussed in detail, it is later introduced). In this regard, the first mention of

LBGR does appear to encourage users to rely on site-specific information for setting the LBGR (at least in Scenario A). It also mentions that for Scenario B the LBGR is the action level (AL) itself. The section refers to Sections 5.3.3.1, 5.3.4.1, Appendix D, Section D.1.7.3 for more information. Note that a reference to Chapter 4, Section 4.12.3.1 WRS Test should perhaps also be included here.

Page 2-12, Lines 25-28 provide more information on how to specify the relative shift that depends heavily on the LBGR in both Scenario A and Scenario B.

Page 2-35, Section 2.5.4, page 2-35, Lines 11-15, notes that the number of measuremenst needed can be recalculated using different values for the LBGR amongst other variables until an appropriate survey design is obtained by the planning team (Appendix M is noted to discuss prospective power curve usage for considering the effects of these parameters).

Page 4-55, Section 4.12.3.1 WRS Test, Lines 22-23 and following, currently states, "The LBGR is often set at the expected median concentration of the radionuclide. However, in our example the mean is higher than the median. Because it is conservative to set the LBGR at the higher value (i.e., the expected mean)…". This is the first section to discuss how to set the LBGR but is does so in an example manner and appears confusing.

Page 4-56, Section 4.12.3.2 Sign Test, Lines 21-23 has a much clearer description, "The LBGR is often set at the expected median concentration of the radionuclide. Because it is conservative to set the LBGR at the higher of the mean or median, the median is used…"

Page 4-58, Section 4.12.4.1 Applying the Unity Rule, Lines 16-18 utilizes the earlier unclear language (same as on Page 4-55) in the example shown. Such inconsistencies lead to confusion for potential users.

It is understood that the definition of the gray region (and hence the LBGR) is discussed in more detail in Section 5.3.3.1. However, while these example discussions in Chapter 4 are helpful it would be much clearer if Rev. 2 would include a short paragraph specifically on 'Considerations for Setting the LBGR' prior to these examples, that discusses how the site-specific information along with some emphasis on conservatism is applied to this most important parameter.

Pages 5-28 and 5-29, Section 5.3.3 Determining Numbers of Data Points for Statistical Tests for Residual Radioactive Material Present in Background, Section 5.3.3.1 Define the Gray Region, includes some discussion of the LBGR, specifically, "For Scenario A, the LBGR is typically chosen to represent a conservative (slightly higher) estimate of the mean concentration of residual radioactive material remaining in the survey unit at the beginning of the FSS. If there is no information with which to estimate the residual radioactive material concentration remaining, the LBGR may be initially set to equal one-half of the $DCGL_W$." It should be noted that later, in the discussion of Scenario B, there is no description of how LBGR is to be set, it only states that "In Scenario B, the UBGR is equal to the DL, and the LBGR is equal to the AL." The user is left to perhaps assume that the discussion on how to select an LBGR for Scenario A could (or should?) be applied for Scenario B and then to set the AL equal to the LBGR.

Pages 5-33 and following, Section 5.3.4 Determining Numbers of Data Points for Statistical Tests for Residual Radioactive Material Not Present in Background, Section 5.3.4.1 Define the Gray Region, also does not discuss specifically how to choose LBGR for Scenario B, although it aptly points out that "In Scenario B, the LBGR is equal to zero or the $DCGL_W$, and the UBGR is defined as the DL." Much later in MARSSIM Rev 2, when discussing interpretation of survey results (Page 8-20, Lines 28-29) it is stated that "Under Scenario B, when there are too many measurements or samples with concentrations above the LBGR, we reject the null hypothesis that the survey unit does meet the release criteria." This is better plain language description and should perhaps be utilized in Chapters 4 and 5 when discussing the LBGR.

MARSSIM Rev 2 does call out NUREG-1505 (NRC 1998a) as a useful reference for more information. The NUREG-1505 includes much clearer language on the site-specific considerations for selecting the LBGR, "…the concentration level below which further remediation is not reasonably achievable" (NUREG-1505, page 3-13).

3.2. Please comment on whether avoiding the use of the term "area factor" improves understandability of the elevated measurement comparison concept (Section 8.6.1).

Area factors, which are simply the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria, should be based on site-specific modeling or calculations. Due to the misapplication of published area factors from the literature and to provide focus on the need for development of site-specific EMC criteria, MARSSIM, Revision 2 avoids the use of the term area factor. In addition, lessons learned from training MARSSIM show that describing the EMC concept in descriptive language, rather than by defining additional terminology, seems to improve understandability of the concept.

Response:

The term 'Area Factor' appears to still be utilized in many sections of the MARSSIM, Rev 2, draft. As such, it is not clear if avoiding the use of the term has been complete or would improve the understandability. In fact, the sections that do still refer to an area factor may introduce significant confusion more generally. A few examples from Rev 2 draft are discussed below to illustrate potential problems introduced by this attempt:

- Section 5.3.5.1, paragraph on page 5-37, Lines 15-22,

> Revisions 0 and 1 of MARSSIM (published in 1998 and 2000, respectively) included the calculation of an area factor[7] as an intermediate step in the determination of the required scan MDC. The use of an area factor is not necessary if $DCGL_{EMC}$ is tabulated directly as a function of the area of radioactive material. To simplify the determination of the required scan MDC, the use of the area factor as an intermediate calculation is not included in this revision of MARSSIM. The area factor can still be used if the ratio of the $DCGL_{EMC}$ to the $DCGL_{W}$ is known and will produce the same results as the approach described in the current revision of MARSSIM.

In section 5.3.5.1 *Determination if Additional Data Points are Needed*, a paragraph notes that revisions 0 and 1 of MARSSIM included the calculation of an area factor (note that it is defined, identified as A_m, and discussed further in footnotes and in symbols list, and specifically given much description on Appendix O, Section O.4) as an intermediate step in the determination of the required scan MDC. It then states that the use of an area factor is not necessary if $DCGL_{EMC}$ is tabulated directly as a function of the area of the radioactive material. This statement could be interpreted to mean that an area factor is needed if $DCGL_{EMC}$ is not tabulated in this manner. However, the section further notes that "the use of the area factor as an intermediate calculation is not included in this revision of MARSSIM". But then it further states that the area factor can still be used under certain conditions. This seems patently confusing. The message appears to be, we don't want you to use area factors, but you can if you want to under certain circumstances.

- Section 5.3.8, paragraph on page 5-49, Lines 3-12,

> When an investigation level is exceeded, the first step is to confirm that the initial measurement or sample actually exceeds the particular investigation level. This may involve taking further measurements to determine that the area and level of the elevated residual radioactive material are such that the resulting dose or risk meets the release criteria. Rather than—or in addition to—taking further measurements, the investigation may involve assessing the adequacy of the exposure pathway model used to obtain the DCGLs and area factors, as well as the consistency of the results obtained with the HSA and the scoping, characterization, and RAS surveys. Depending on the results of the investigation actions, the survey unit may require reclassification, remediation, or resurvey. **Table 5.4** illustrates an example of how investigation levels can be developed.

In Section 5.3.8, Determining Investigation Levels, a paragraph discussing steps to take when an investigation level is exceeded, points out that rather than – or in addition to – taking further measurements, the investigation may involve assessing the adequacy of the exposure pathway model used to obtain the DCGLs and *area factors*, etc. This is confusing, as section 5.3.5.1 has already noted that area factors are not necessary for Rev 2. Perhaps delete the reference to area factors here or refer to the ratio approach.

- Section 8.6.3, example 14 box on page 8-49.

> **Example 14: Class 1 Survey Unit with Elevated Areas**
>
> Consider a Class 1 Survey unit that passes the nonparametric statistical tests and contains some areas that were flagged for investigation during scanning. Further investigation, sampling, and analysis indicate one area is truly elevated. This area has a concentration that exceeds the derived concentration guideline level determined using the Wilcoxon Rank Sum test by a factor greater than the area factor calculated for its actual size. This area is then remediated. Remediation control sampling shows that the residual radioactive material was removed, and no other areas were affected by residual radioactive material. In this case, one may simply document the original final status survey (FSS), the fact that remediation was performed, the results of the remedial action support survey, and the additional remediation data. In some cases, additional FSS data may not be needed to meet the release criteria.

This example describes ‘one area is truly elevated’ noting that “the area has a concentration that exceeds the derived concentration guideline level determined using the Wilcoxon Rank Sum test by a factor greater than the *area factor* calculated for its actual size.” Utilizing the term ‘area factor’ in such an example appears to introduce confusion as it appears as an important deciding factor in describing if an area is truly elevated. However, Section 5. 3.5.1 has already noted that area factors are not necessary for Rev. 2. This example should be re-written to remove this confusion.

- Appendix O, Section O.4 on pages O-4 through O-8.

This entire section discusses ‘Calculating Area Factors and the DCGL for the EMC’. While it notes that ‘Because these area factors were misused for specific problems, the term “area factor is largely omitted from the main body of this report. Historical information on the use of area factors is provided in this appendix for completeness". This is very confusing to a reader. If this important line is not well understood, users might see the illustrative examples and Table O.4 (Illustrative Examples of Outdoor Area Factors) and Table O.5 (Illustrative Examples of Indoor Area Factors) as able to be used. It is recognized that Rev. 2 does include footnotes noting these are illustrative only, but these footnotes also direct users to 'Consult regulatory guidance to determine area factors to be used for compliance demonstration.' This leads us right back to the issue noted in Section 5.3.5.1 that has already noted that area factors are not necessary. These footnotes seem to say that they are so necessary that users should consult with their regulator. Perhaps better to not include any discussion about area factors if not needed. Why address historically in this appendix at all? Further, the summary and conclusions statements in Section 0.4.5 note that "It is always acceptable and conservative to assume the smallest area factors possible (i.e., 1)." If area factors are not necessary, then this conclusion is highly confusing to a user.

- Appendices – Glossary, page GL-1, includes a definition of area factor, but does not state using it is not necessary. Perhaps this should refer to a ratio of $DCGL_{EMC}/DCGL_W$ or not be included in the glossary at all.

> **area factor (A_m):** A factor used to adjust the *derived concentration guideline level ($DCGL_W$)* to estimate the *derived concentration guideline level ($DCGL_{EMC}$)* and the *minimum detectable concentration* for *scanning surveys* in *Class 1 survey* units, wherein the $DCGL_{EMC} = DCGL_W \times A_m$. A_m is the magnitude by which the *concentration* of *residual radioactive material* in a small *area of elevated activity* can exceed the $DCGL_W$ while maintaining compliance with the *release criteria*.

3.3. Please comment on the effectiveness of the new organization of Chapter 4 (Considerations for Planning Surveys) to improve the understandability of the Chapter.

Earlier reviews of Chapter 4 provided evidence that the fundamental organization of Chapter 4 made it difficult to find and understand vital information. After discussing the challenge with experts in training and explaining the material, Chapter 4 was completely rewritten or reorganized in an attempt to improve understandability without changing the fundamental purpose of or material in the Chapter. In an effort to streamline the presentation of material in Chapter 4, some information was moved to Appendix O.

Response:

MARSSIM Rev 1 (2000) Chapter 4 was certainly less organized and did make it difficult to find and understand vital information, mostly because it didn't follow the flow of the survey process clearly. MARSSIM Rev 2 (draft) Chapter 4 is much more organized and follows the flow of the survey process more clearly, aiding in overall understandability. It would be helpful for Chapter 4 to perhaps refer to the flow charts in Chapter 2 (Figures 2.4 through 2.8) to remind users of the overall process applicable in each of the sub sections in Chapter 4. The callouts from Chapter 4 to other subsequent Chapters (e.g., Chapter 5 for more information on survey types, and the Glossary for specific definitions) should also be very helpful for users. Flowcharts in Figures 4.1 and Figures 4.2 are particularly useful (along with the Section callouts). Moving the derivations to Appendix O, Detailed Calculations for Statistical Tests and Illustrative Examples for the Determination of DCGLS, is useful and helps maintain the flow of the material presented in Chapter 4. The Sections in Chapter 4 that include "Notes" or "Important Considerations" will likely be very useful for users and should help avoid problems downstream during regulatory review. The "Conclusions" at the end of each of the foregoing discussions is likely very helpful for users. Perhaps more of them should be included as endings of each major Section in Chapter 4.

Page 4-25, Lines 6-25, Section 4.7.1 Basic Terms, should perhaps include callouts to other MARSSIM Rev 2 Sections where these concepts are specifically addressed with more detail, or should provide footnotes with definitions, glossary entries, or reference to other materials, rather than just leave them to the user to either lookup (outside of MARSSIM process) or find (within the MARSSIM document).

3.4. Please comment on the effectiveness of moving derivations from Chapter 5 to Appendix O to improve the understandability of the Chapter.

In an effort to streamline the presentation of material in Chapter 5, some derivations of key concepts were moved to Appendix O.

Response:

Moving the derivations to Appendix O, Detailed Calculations for Statistical Tests and Illustrative Examples for the Determination of DCGLS, is useful and helps maintain the flow of the material presented in Chapter 5 as well. However, Section 5.1 Introduction should specifically include a sentence callout to Appendix O. It is currently only identified in discussions or footnotes associated with Tables 5.2 and 5.3. Perhaps something like, "Appendix O provides detailed calculations for statistical tests, illustrative examples for the determination of DCGLs, and more detailed derivations of key statistical concepts and should be consulted when undergoing survey planning and design."

Dr. Timothy A. DeVol

Draft Preliminary comments on the assigned charge questions

Charge Question 1.1 (Associate Reviewer)
"Please identify whether the inclusion and proposed implementation of scan-only surveys (Section 5.3.6.1 and Section 8.5) is appropriate, adequate and clear, especially the discussion on sampling for scan-only measurement method validation or verification."

I feel that the current description of scan-only surveys is appropriate, adequate and clear; however, I feel that it will continue to be misinterpreted. The title, "scan-only" puts one in the mind-set that they can just scan and go. If the title "Scan-Only Surveys" was changed to "Surveys with site-specific validated scan system", I feel that the misinterpretation would be minimized.

Specific Comments

- There is no reference to Table 5.5 which summarizes Section 5.3.6.1
- Page 5-42, line 25 - Recommend specifying the scan coverage from 10-100% rather than the use of the vague statement of "a much larger portion". The use of "much larger" is subject to interpretation.
- Page 5-43, line 7 – It is specified that the scan MDC should be less than 50% of the $DCGL_W$. What is the basis for the stated 50% reduction?
- Page 5-43, line 8 – It is specified that "some percentage" of the samples go for laboratory analysis. It is now explicit that the "scan-only" has to be accompanied with sampling and laboratory analyses. I feel that this is the reason for the misinterpretation; in Section 5.3.6.2 is titled "Scanning and Sampling". One need to read further into Section 5.3.6.2 to sort out the differences.
- Page 5-43, Equation 5-10 – It would be more accurate if "Scan Area" were relabeled "% of Scan Area".
- Page 5-43, lines 21-22 – It seems to me that the issue being raised here is applicable to more than just alpha and beta radiation. It seems that the issue is the assumption of whether there is surface or volumetric contamination at the scan survey site. The results will be impacted for alpha, beta and gamma-ray emitting radioactive materials, but will have a greater impact on alpha, beta, and low-energy gamma-ray and x-ray radiation.

Charge Question 1.4 (Associate Reviewer)
"*Is the discussion of survey requirements for areas of elevated activity technically accurate, appropriate and clear? In particular, please comment on the decision to maintain the use of the unity rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4).*"

Personally, I found the presentation in MARSSIM Rev. 2 on this topic very confusing. I am reading this section for the first time, and not reading the material leading up to Section 5.3.5. Appendix O.4 is helpful in understanding the background on the elevated activity and the advantages/disadvantages of the different approaches. The examples are helpful in understanding the process that is to be followed, but the notation is inconsistent making the equations leading up to the examples hard to follow.

Specific Comments

- Page 5-36, line 3 – Should read "…treatment of areas of elevated radioactive materials…". The adjective "elevated" is misplaced.
- Page 5-36, line 30 – It would be helpful to the reader to give a little insight into the decision on selection of a triangular or rectangular grid. There are certainly advantages/disadvantages, which should be briefly presented before sending the reader off to EPA 1994b for a detailed explanation.
- Page 5-36, line 31, Equations 5-1 and 5-2 – The notation in Section 5.3.5 is inconsistent and confusing. For example, *A* is defined as the "total area of the survey unit". But in Equations 5-1 and 5-2, the total area of the survey unit now seems to be defined as "*A* (survey unit)"
- Page 5-37, Equations 5-3 and 5-4 – "*A* (grid area)" is used in these equations and never defined. Further "*A* (grid area)" and "*A* (surface area)" are not defined in "Symbols, Nomenclature, and Notations," Page xxviii. What is defined on Page xxviii for "*A*" is overall sensitivity of a measurement, whereas "A" is area.
- Page 5-37, line 4 – "A_{EA}" is defined here and is consistent with "Symbols, Nomenclature, and Notations" on Page xxviii as area of elevated activity. I point to this as an example of good and logical notation that should be used consistently throughout Section 5.3.2 and the entire MARSSIM document.
- Page 5-37, Equations 5-5 and 5-6 – The use of "Scan MDC (actual)" and "Scan MDC (required)" is confusing in these two equations and never defined. If someone jumps to these equations and starts to apply then without reading the context for which they apply, they will come to the conclusion that Scan MDC (actual) = Scan MDC (required). To avoid the possible confusion and the poor notation I recommend that these two equations be removed.

Charge question 2.1 (Lead Reviewer)
"Please comment on the usefulness and accuracy of updated measurement methods and instrumentation information (Chapter 6 and Appendix H)."

I found the updated measurement methods and instrumentation information useful and for the most part accurate. However, there exists opportunities to make the information even more useful. Although MARSSIM should not be an advertising platform, it would be useful to have example manufacturer's name and model number for the instrumentation that is being discussed.

If it is not appropriate for the individual instruments, then present this information generically in a table so that "*…those interested in purchasing or using the equipment are encouraged to contact vendors …*" have an idea where to go. Further, there is specific information in Appendix H description that should be properly referenced, for example Page H-10, lines 23-25. The statement was clearly the result of some literature that should be properly cited so the reader can go obtain additional information. Also by including references to the scientific literature information of the manufacturers and models would be available to the reader thus getting around the issue of being a potential advertising platform.

Appendix H needs an extensive technical edit. For the most part, the descriptions of the individual instruments is good, but when compared with the other instruments there are inconsistencies. Further there are things that are stated that just seem incorrect, for example Page H-2, lines 19-20, "*That allows probes to transmit data signal to the receptors, such as BT (Bluetooth)- equipped computers, phones or planchets*". I believe that the authors meant to state "tablets" rather than "planchets".

<u>Select</u> Specific Comments

- Page 6-2, line 32 – Recommend wording to be "background of the specific radionuclides of interest". "Specific" should modify radionuclides, not background.
- Page 6-35, lines 28, 29 – Consider replacing "neon or helium" with "noble gas" that is more inclusive to the possible gasses used. Also consider removing the reference to "quenching agent". Although it is correct, methane (lines 26-27) is also referred to as a "quench agent", but the two quench agents are quenching different phenomenon. To avoid that level of detail, it would be easiest to remove the reference to the quench agent.
- Page 6-35, lines 35, 36 – Use of the descriptor "activator" is not used consistently in Chapter 6 and Appendix H when referring to scintillators, and TLD and OSL materials. One example is being highlighted here, but there are numerous. NaI(Tl) should be referred to as "thallium-activated sodium iodide". The use of " -activated" is not used consistently in these sections of the document.
- Table 6.8 – There are advantages and disadvantages to instruments and there are advantages and disadvantages to measurement technique; it becomes very repetitive to type to address both in a single table. The summary table is good, but may be better if it were two tables.
- Page 6-49, Table 6.8 – The use of "Laboratory Analysis" as the "Instrument" is not appropriate.
- Page 6-53, line 14 – The reference to Jenkins 1986 leads the reader to believe that there have been no new developments in the last 3+ decades. Consider updating the reference.
- Table 6.9 – The column "Time" is not clear; is that "total time", "sample collection time" or "count time"
- Table 6.9 – Regarding the "Remarks" column, "LLD" should be replaced with "MDC". Also the MDC listed are in a variety of different units which makes it difficult for the

reader to compare/contrast the different techniques quickly and easily. Consider listing the MDC is the same units for all methods. This table is another example where the use of citations to the scientific literature would greatly enhance the information contained in the table. With the citation, the reader has the option to quickly and easily getting more information on the method.

- Page H-12, lines 42, 43 – It appears that some of the cost estimates of the instruments have not been updated since Revision 1 (2000). This is a specific example of a potential discrepancy, but there are others. The cost estimates on all the instruments should be updated to 2020 dollars.
- Page H-14, line 1 – Should include CdTe detectors. Both CdTe and CZT are commercially available and utilized in research and COTS instruments.
- Page H-14, lines 36, 37 – Clarification should be added to these costs to indicate that this is the cost of the detector without the associated data acquisition system. Should follow the example for the PIC (Page H-10, lines 34-37). Want to strive for consistency in the reporting of the costs. This is one example of the discrepancy of reporting. There are other incidents. This becomes more of an issue when presented in Tables H.2 – H.8 where costs are presented for comparison where they are not equivalent.
- Appendix H – It seems appropriate to include "Direct Ion Storage" (DIS) devices in Appendix H. TLDs, OSLs, EDs are all presented, but DIS is not. DIS devices are commercially available and are "drop in" replacements for TLD and OSL. I feel that it is appropriate to be presented in Appendix H.

Charge Question 2.2 (Associate Reviewer)
"Please comment on the usefulness and accuracy of the additional optional methodology for the use of Ranked Set Sampling for hard-to-detect radionuclides."

I feel that the Appendix E is useful for implementation of the RSS methodology. Although I am not a statistician and was not able to follow all the methodology, the description with examples were helpful in following the methodology. It is stated (Page E-1, lines 40-42), "*Performing an RSS survey requires a much greater level of expertise in survey planning and implementation than a traditional Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) survey requires.*" A reference is given to give the reader more depth into the topic.

Specific Comments

- Page E-1, line 40 through page E-2, Line 2 – Recommend moving this paragraph to the beginning of Appendix E so that it is immediately available to the reader.
- Page E-6, line 6 – Need to insert a space between "median" and "of".
- Page E-12, lines 1,2, 4 – The line numbering in the document should be removed

- Page E-14, Example 3 – There is a change in notation in this example. Initially the mean is defined with the symbol "μ_{RSS}" then when calculating sample variance the mean is defined as "$\overline{X}_{RSS}$". Should use consistent notation to avoid confusion
- Page E-14, lines 3-9 – The line numbering in the document should be removed

Mr. Earl Fordham

1.4. Is the discussion of survey requirements for areas of elevated activity technically accurate, appropriate and clear? In particular, please comment on the decision to maintain the use of the unity rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4). **Are there suggested alternatives to the use of the unity rule?**

While modeling is outside the scope of MARSSIM, depending on the modeling tool or methodology used to develop release criteria, the use of the Unity Rule for multiple areas of elevated activity in a single survey unit can lead to unrealistic or overly conservative assumptions. For example, the models may assume that the receptor is located directly above each area of elevated activity and stays there for the duration of their exposure period. This physically cannot occur in cases where there is more than one area of elevated activity per survey unit and results in concerns that this will cause an over-estimate of dose or risk, leading to an emphasis on remediating areas of elevated activity that don't incur additional significant dose or risk to receptors.

MARSSIM, Revision 2, does not change recommendations for the use of the unity rule, but emphasizes assessing whether criteria for areas of elevated activity apply to survey units, and when they do, using a common sense approach to applying these criteria, keeping in mind the limitations of the unity rule described above for multiple areas of elevated activity.

Response: In the narrative in Chapter 8 Section 6, use of the unity rule is suggested to ensure the total dose is within the release criteria. Using the unity rule for a single elevated primary area is a good selection, however, the unity rule for two or more elevated areas has a potential to cause an overly conservative (and potentially impossible) scenario. These types of scenarios lend themselves to alternative approaches, which are alluded to in the text. The user needs to be well versed in the modeling software being used to develop DCGL's. When defaults are no longer properly used (e.g., resident time as a percentage at a location), the user need to discuss the issue and obtain concurrence from the regulatory body overseeing the remediation and release of the building and/or grounds.

Using the unity rule to ensure total dose is within the release criteria can cause unnecessary effort and expense. As shown in Abelquist (2008), doses from elevated areas other than a single primary area can be relatively small and negligible. Additional consideration from a second (or additional) elevated area may require changes to the underlying exposure scenario (e.g., percentage time at a given location). To accomplish this change, the regulatory agency will need to notified, briefed and concurrence obtained, which may not be cost effective. I suspect in most cases, early remediation of multiple elevated areas will circumvent the need to use a complex unity formula with additional terms addressing individual elevated areas.

1.5. Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles **technically sound and appropriate, and is the description accurate**? In particular, please comment on the rule-of-thumb for determining when use of MARSSIM may not be appropriate for survey units containing discrete radioactive particles (Section 4.12.8 and Appendix O.5).

Discrete radioactive particles have an extremely small size and contain enough activity that survey units containing discrete radioactive particles generate impractical survey designs under MARSSIM. Over MARSSIM's twenty-year history, several sites have attempted to utilize MARSSIM to address discrete radioactive particles, with predictably extreme survey designs as a result. In addition to being impractical, designs for discrete radioactive particles violate some of the assumptions commonly made during modeling, which includes parameters based on an areal source of radioactive material, e.g., length of the area of the elevated activity in the direction of overland flow. While modeling is outside of the scope of MARSSIM, it is nonetheless required that survey designs match the assumptions made during modeling, otherwise, the survey design does not meet the requirements of the action level.

To set a limit for determining when areas of elevated activity are too small to use the traditional MARSSIM methodology, the MARSSIM Workgroup used a traditional rule-of-thumb for instrumentation. When the length of the area of elevated activity is less than three times the distance to the detector, the area of elevated activity is viewed by the detector as a point source instead of as an areal source. These point sources will need different receptor modeling and release requirements, and hence different survey designs than traditional areal sources.

At this time, MARSSIM does not provide guidance on designing discrete radioactive material surveys. It is the intention of the revision that additional information provided should prevent MARSSIM from being applied inappropriately to survey units involving discrete radioactive particles.

Response: The use of the rule-of-thumb (d > 3l) is proper for determining when a spot of contamination should be considered a point source or a plane source. When isolated discrete radioactive particles (DRP's) are located during a survey (e.g., scan with portable instruments), the initial assumption should be that the instrument to source distance (d) will be much larger than three times the longest length (3l) of the source thus resulting in the DRP being treated as a point source. The microscopic (typical) size of the DRP is the controlling factor.

DRP's (or "hot particles) contain a very high specific activity. Trying to design a realistic survey scenario with DRP's scattered in the survey unit is simply not beneficial. The survey parameters would drive the size of the survey unit to a very small area that a

MARSSIM Final Status Survey (FSS) would not result in any cost savings which is a hallmark of MARSSIM. Since the FSS is by design the last survey to be completed at a location, it would be more appropriate that earlier surveys (e.g., scanning with instruments mounted on bumpers of trucks) identify the DRP's and remedial activities (i.e., RAS survey) remove the DRP's before a FSS is conducted.

3.2. Please comment on whether avoiding the use of the term "area factor" improves understandability of the elevated measurement comparison concept (Section 8.6.1).

Area factors, which are simply the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria, should be based on site-specific modeling or calculations. Due to the misapplication of published area factors from the literature and to provide focus on the need for development of site-specific EMC criteria, MARSSIM, Revision 2 avoids the use of the term area factor. In addition, lessons learned from training MARSSIM show that describing the EMC concept in descriptive language, rather than by defining additional terminology, seems to improve understandability of the concept.

Response: Concur, I believe the understanding of Section 8.6.1 is improved without the term "area factor". While the narrative in this section provides necessary detail, I would suggest a figure or two be added to enhance the reader's understanding.

3.4. Please comment on the effectiveness of moving derivations from Chapter 5 to Appendix O to improve the understandability of the Chapter.

In an effort to streamline the presentation of material in Chapter 5, some derivations of key concepts were moved to Appendix O.

Response: Moving derivation and underlying reason from the body of the report to the appropriate Appendix enhances the comprehension and readability of the document and should be done wherever possible.

Dr. Eric M Goldin

1.1 Please identify whether the inclusion and proposed implementation of scan-only surveys (Section 5.3.6.1 and Section 8.5) is appropriate, adequate and clear, especially the discussion on sampling for scan-only measurement method validation or verification.

> The MARSSIM Workgroup wrote MARSSIM, Revision 1, for 1995 technology, not envisioning that future instrumentation would be able to measure a statistically significant portion of the survey unit while meeting required Measurement Quality Objectives (MQOs), especially that the Minimum Detectable Concentration (MDC)/Minimum Detectable Activity (MDA) be less than 50% of the Derived Concentration Guidelines Level for wide areas (DCGLw). New methods for designing, implementing and assessing scan-only surveys are included in the revisions to make effective use of resources when employing these technologies.
>
> Earlier reviewers misinterpreted the term “scan-only surveys” to mean that samples would not be taken as any part of the survey process. Revision 2 has been further revised to indicate that quality control samples may need to be collected as part of the method validation or verification process, as appropriate.

The inclusion and implementation process for scan-only surveys is appropriate and definitely needed (Section 5.3.6.1). The discussion is adequate and clear, and is a necessity for many sites that have a condition where the licensee is being driven (by either regulatory authority or stakeholder pressure) to achieve “nothing detectable above background.”

I particularly liked the description in Section 5.3.6 that explains when a scan-only survey can be implemented compared to a scan survey plus sampling. The comparison of measurement capability to instrument sensitivity is clear and important.

The inclusion of a description of advanced technology that was lacking in MARSSIM Rev. 1 is excellent and greatly needed.

The inclusion of the quality control sample process is also appropriate, well-explained, and necessary. The process for determining the percentage of area to be scanned is also clear and easy to understand.

Section 8.5, on the other hand, is not as clear and easy to follow. The description of the complicated statistical analysis could have been improved with an example. The section ends with a cautionary statement but not enough detail to help the reader understand what are the risks from a small data set. Again, an example might help present this concept in a more understandable form.

1.5. Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles technically accurate, appropriate and clear? In particular, please comment on the rule-of-thumb for determining when use of MARSSIM may not be appropriate for survey units containing discrete radioactive particles (Section 4.12.8 and Appendix O.5).

> Discrete radioactive particles have an extremely small size and contain enough activity that survey units containing discrete radioactive particles generate impractical survey designs under MARSSIM. Over MARSSIM's twenty-year history, several sites have attempted to utilize MARSSIM to address discrete radioactive particles, with predictably extreme survey designs as a result. In addition to being impractical, designs for discrete radioactive particles violate some of the assumptions commonly made during modeling, which includes parameters based on an areal source of radioactive material, e.g., length of the area of the elevated activity in the direction of overland flow. While modeling is outside of the scope of MARSSIM, it is nonetheless required that survey designs match the assumptions made during modeling, otherwise, the survey design does not meet the requirements of the action level.
>
> To set a limit for determining when areas of elevated activity are too small to use the traditional MARSSIM methodology, the MARSSIM Workgroup used a traditional rule-of thumb for instrumentation. When the length of the area of elevated activity is less than three times the distance to the detector, the area of elevated activity is viewed by the detector as a point source instead of as an areal source. These point sources will need different receptor modeling and release requirements, and hence different survey designs than traditional areal sources.
>
> At this time, MARSSIM does not provide guidance on designing discrete radioactive material surveys. It is the intention of the revision that additional information provided

should prevent MARSSIM from being applied inappropriately to survey units involving discrete radioactive particles.

The discussion on discrete radioactive particles is technically accurate (description is correct), appropriate (needed), and clear (presents the condition, evaluation, and a path to deal with the problem). However, Appendix Section O.5 is a verbatim repetition of Section 4.12.8. It is not clear that the Appendix section is needed if it simply repeats the text.

Perhaps there could have been some discussion related to remediation if hot particles are detected. In other words, a scan survey can detect the particles which then may be removed from the area, and the area resurveyed. A good scan survey should detect discrete radioactive particles as long as the detector to particle distance is not too great. My experience was: a large scan survey was conducted, a discrete particle was discovered, the particle was removed and the area was rescanned. Upon rescanning with additional scanning of "thinner" lifts of soil resulted in no additional hot particles being discovered.

The section on detectability using the rule-of-thumb comparing 3 times the distance-to-detector to the largest dimension of the contaminant is good. It makes sense and is clear. Matches well with some rules-of-thumb regarding point source geometry.

In some sense, the discussion of hot particles could be enhanced by a description of the impact - if hot particles are present, some detail on the type, radiological properties, and dose impacts should be included. Then a detailed discussion of hot particle detection capability would need to be included. Finally, some decision-making should be required that shows that if hot particles are present, and if they are detectable, describe the basis for either decontaminating the area or what the dose impact is if the particles are to be left. In many respects, if particles are present and detectable, they should simply be removed. If it is deemed that no additional text is necessary for the discussion of hot particles, then perhaps a statement should be included that refers the licensee to communicate with the appropriate regulatory agency regarding the potential presence and dose consequence of hot particles.

2.3 Please comment on the usefulness and accuracy of new and additional examples provided in Chapter 5.

The examples that appear in Chapter 5 include:

1. Scoping Survey Checklist, pages 5-7 - 5-9 (same as Rev. 1, pages 5-5 - 5-6)

2. Characterization Survey Checklist, pages 5-17 - 5-19 (same as Rev. 1, pages 5-16 - 5-17)

3. Remedial Action Support Survey Checklist, page 5-21 (same as Rev. 1, page 5-20)

4. Use of WRS Test under Scenario A, page 5-31 (modified from Rev. 1, pages 5-33 - 5-35)

5. Use of WRS Test under Scenario B, page 5-33 (new)

6. Use of Sign Test under Scenario A, page 5-34 (new)

7. Determination Whether Additional Data Points are Required, pages 5-39 - 5-40 (modified from Rev. 1, page 5-39)

8. Determination Whether Additional Data Points are Required, pages 5-40 - 5.42 (modified from Rev. 1, page 5-39)

9. Random Sampling Pattern, page 5-45 (same as Rev. 1, page 5-41)

10. Illustration of a Triangular Systematic Pattern in an Outdoor Class 2 Survey Unit, pages 5-46 - 5-47 (same as Rev. 1, page 5-43)

11. Final Status Survey Checklist, pages 5-56 - 5-58 (modified from Rev. 1, pages 5-53 - 5-55)

Of the 11 examples that appear in Chapter 5, five are unchanged, four are modified (additional information), and two are new.

The new examples are a significant improvement and are needed. The unchanged examples were mostly reformatted such that they are easier to read and understand. The modified examples included some additional information or steps or changes to data such that they are better suited to inform the reader.

In summary, the examples in Chapter 5 are a marked improvement over Rev. 1 examples.

3.4. Please comment on the effectiveness of moving derivations from Chapter 5 to Appendix O to improve the understandability of the Chapter.

> In an effort to streamline the presentation of material in Chapter 5, some derivations of key concepts were moved to Appendix O.

The derivations that were once in Chapter 5 have been moved to Appendix O, sections O.2 (WRS test), O.3 (Sign test), and O.4 (Area Factors and EMC). This is a great improvement over the text in MARSSIM Rev. 1 Chapter 5. However, there should be more references pointing to Appendix O from the appropriate sections in Chapter 5.

For example, Section 5.3.3 presents requirements for completing the WRS but doesn't actually refer the reader to Appendix O, Section O.2 except for a tiny footnote under Table 5.2. Similar situations exist for Section 5.3.4 on the Sign Test (should refer to O.3 beyond the footnote under Table 5.3), and 5.3.5 on the EMC (no reference at all to Appendix O.4).

Dr. Barbara Hamrick

1) Are the revisions to MARSSIM concepts and methodologies **technically appropriate, useful and clear**, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

1.1 Please identify whether the inclusion and proposed implementation of scan-only surveys (Section 5.3.6.1 and Section 8.5) is appropriate, adequate and clear, especially the discussion on sampling for scan-only measurement method validation or verification.

> The MARSSIM Workgroup wrote MARSSIM, Revision 1, for 1995 technology, not envisioning that future instrumentation would be able to measure a statistically significant portion of the survey unit while meeting required Measurement Quality Objectives (MQOs), especially that the Minimum Detectable Concentration (MDC)/Minimum Detectable Activity (MDA) be less than 50% of the Derived Concentration Guidelines Level for wide areas (DCGLw). New methods for designing, implementing and assessing scan-only surveys are included in the revisions to make effective use of resources when employing these technologies.
>
> Earlier reviewers misinterpreted the term “scan-only surveys” to mean that samples would not be taken as any part of the survey process. Revision 2 has been further revised to indicate that quality control samples may need to be collected as part of the method validation or verification process, as appropriate.

RESPONSE: The inclusion and proposed implementation of scan-only surveys is appropriate. I am not convinced the discussion of it is adequate and clear. As I read Section 5.3.6.1, it says validation is required, but also says that validation will only “*likely* require site-specific…samples” (emphasis added). The implication being that there may be other ways to validate the scan-only results. If that is the intent (i.e., that the scan-only results may be validated in some way that does not involve taking site-specific samples), then I think that should be expressly stated. If that is not the intent, then “likely” and “may” should be removed as qualifiers with respect to the need for site-specific validation samples.

1.2 Please comment on the inclusion and proposed implementation of Scenario B (Chapter 4, Section 5.3, and Chapter 8). Is it appropriate to recommend that Scenario B be used only for those situations where Scenario A is not feasible? Are methods for considering background variability in assessing whether the site is indistinguishable from background reasonable and technically accurate? Is the inclusion and proposed implementation of added requirements for retrospective power analysis and the Quantile Test while using Scenario B technically appropriate and discussed adequately and clearly?

> Under hypothesis testing in MARSSIM, Scenario B is defined as assuming that the

survey unit meets the release criteria unless proven otherwise, and its use was discouraged in MARSSIM, Revision 1. However, this is the only viable option for sites where the criterion is effectively “no added radioactivity” or “indistinguishable from background”.

In Scenario B, the Lower Bound of the Grey Region (LBGR) is often set to zero, but the document allows use of a non-zero LBGR that considers background variability in determining whether the survey unit is indistinguishable from background.

Since Scenario B assumes that the site meets the release criteria, there is a risk that the survey unit will pass simply because the survey did not have sufficient rigor. To guard against that, the revisions require that when using Scenario B, the survey unit must perform a retrospective power analysis to prove the survey has sufficient statistical power to detect a survey unit that should not have passed.

The non-parametric tests included in MARSSIM test the median instead of the mean. The
release criteria are typically expressed as the mean. To guard against Scenario B situations where the median will pass but the mean won’t (this can occur in sample data distributions with a long tail in the higher concentration range), Revision 2 also requires that when using Scenario B, the survey unit must pass a quantile test to guard against excessive skewness.

RESPONSE: It is appropriate to recommend Scenario B be used for those situations where Scenario A is not feasible, but also, with the addition of the retrospective power analysis and the Quantile Test, I believe it is appropriate to allow Scenario B as an alternative even where Scenario A is not impossible.

Also, I note the language on page 4-51, Lines 10-14 is as follows: “Historically, MARSSIM recommended the use of Scenario A, which put the burden of proof that the survey unit met the release criteria on the individuals designing the survey. In Scenario B, the burden of proof is no longer on the individuals designing the survey and thus should be used with caution and only in those situations where Scenario A is not an effective alternative.”

That language strikes me as carrying unintended implications about “the individuals designing the survey.” The reality is there are, as you point out, cases where Scenario A simply cannot work, and Scenario B is a scientifically defensible alternative. I would rephrase the above to say: “MARSSIM generally recommends the use of Scenario A; however, there are certain situations in which Scenario A cannot be applied (such as those where the release criterion is “indistinguishable from background”), and in those cases MARSSIM recommends the use of Scenario B.

Page 4-51, Lines 18-19, I recommend removing the following sentence: “The choice of Scenario A or Scenario B should be based on which null hypothesis is easier to live with if false (NRC 1998a).” It does not add anything to the discussion.

Page 5-26, Lines 18-30, I suggest you remove the "burden of proof" language. I understand what you mean in terms of selecting the direction of the arrow in the null hypothesis (x > release criterion vs. x < release criterion), but quite seriously, the "burden of proof" is *still upon the people designing the surveys* under either scenario to give confidence to the regulator that the criterion is met. By adding the retrospective power analysis and the Quantile Test, there is adequate support to reason that it is not likely the null hypothesis is in fact false in Scenario B.

The methods for considering background variability in the circumstances given in the question are reasonable and technically accurate. The inclusion and implementation of the retrospective power analysis and the Quantile Test for Scenario B are technically appropriate and discussed adequately and clearly.

1.4. Is the discussion of survey requirements for areas of elevated activity technically accurate, appropriate and clear? In particular, please comment on the decision to maintain the use of the unity rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4). **Are there suggested alternatives to the use of the unity rule?**

> While modeling is outside the scope of MARSSIM, depending on the modeling tool or methodology used to develop release criteria, the use of the Unity Rule for multiple areas of elevated activity in a single survey unit can lead to unrealistic or overly conservative assumptions. For example, the models may assume that the receptor is located directly above each area of elevated activity and stays there for the duration of their exposure period. This physically cannot occur in cases where there is more than one area of elevated activity per survey unit and results in concerns that this will cause an over-estimate of dose or risk, leading to an emphasis on remediating areas of elevated activity that don't incur additional significant dose or risk to receptors.
>
> MARSSIM, Revision 2, does not change recommendations for the use of the unity rule, but emphasizes assessing whether criteria for areas of elevated activity apply to survey units, and when they do, using a common sense approach to applying these criteria, keeping in mind the limitations of the unity rule described above for multiple areas of elevated activity.

RESPONSE: With respect to the discussion of areas of elevated activity, please refer to my response to Charge 3.2. Strictly with respect to the discussion of the unity rule, I think it is fine to retain that with the emphasis (as noted) on the potential for it to cause an over-estimate of dose or risk. The unity rule is a tool for simplifying the assessment, and is useful for that purpose. If it results in a gross over-estimate, then a more refined tool can be used.

1.5. Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles **technically sound and appropriate, and is the description accurate**? In particular, please comment on the rule-of-thumb for determining when use of MARSSIM may not be appropriate for survey units containing discrete radioactive particles (Section

4.12.8 and Appendix O.5).

> Discrete radioactive particles have an extremely small size and contain enough activity that survey units containing discrete radioactive particles generate impractical survey designs under MARSSIM. Over MARSSIM's twenty-year history, several sites have attempted to utilize MARSSIM to address discrete radioactive particles, with predictably extreme survey designs as a result. In addition to being impractical, designs for discrete radioactive particles violate some of the assumptions commonly made during modeling, which includes parameters based on an areal source of radioactive material, e.g., length of the area of the elevated activity in the direction of overland flow. While modeling is outside of the scope of MARSSIM, it is nonetheless required that survey designs match the assumptions made during modeling, otherwise, the survey design does not meet the requirements of the action level.
>
> To set a limit for determining when areas of elevated activity are too small to use the traditional MARSSIM methodology, the MARSSIM Workgroup used a traditional rule-of- thumb for instrumentation. When the length of the area of elevated activity is less than three times the distance to the detector, the area of elevated activity is viewed by the detector as a point source instead of as an areal source. These point sources will need different receptor modeling and release requirements, and hence different survey designs than traditional areal sources.
>
> At this time, MARSSIM does not provide guidance on designing discrete radioactive material surveys. It is the intention of the revision that additional information provided should prevent MARSSIM from being applied inappropriately to survey units involving discrete radioactive particles.

RESPONSE: Section 4.12.8 and O.5 contain exactly the same language. I am not sure repeating it adds any value.

The title of the section, "Release Criteria for Discrete Radioactive Particles" is a misnomer, as it emphatically does not provide any release criteria for such.

There doesn't appear to be any guidance on recognition that hot particles may be an issue (although one theoretically could anticipate them with a thorough historical site assessment). Without having some advance knowledge (e.g., from historical events) that hot particles may be an issue at the site, the rule of thumb is not very useful; nevertheless, it is helpful to have some limitation on when the use of the MARSSIM EMC process is not appropriate, and this rule of thumb is technically sound.

From a practical standpoint, however, I think it might be useful to provide guidance on recognizing that one is dealing with a hot particle, or very small area-high activity deposition. For example, if a single sample or survey result is some multiple of the $DCGL_W$, then it should raise suspicion of a small discrete source.

Perhaps something on discrete particles could be added to Section 5.3.8, "Determining Investigation Levels," and Section 8.6.1, "Elevated Measurement Comparison."

3) Is the information in MARSSIM, Revision 2 clear, understandable and presented in a logical sequence? How can the presentation and content of material be modified to improve the understandability of the manual?

historical site information, or a scoping or characterization survey if the survey unit is un- remediated, or the remedial action survey if the site has been remediated. The purpose of the revisions is to describe this concept in plain language, moving away from a statistics terminology description of the concept.

3.2. Please comment on whether avoiding the use of the term "area factor" improves understandability of the elevated measurement comparison concept (Section 8.6.1).

Area factors, which are simply the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria, should be based on site-specific modeling or calculations. Due to the misapplication of published area factors from the literature and to provide focus on the need for development of site-specific EMC criteria, MARSSIM, Revision 2 avoids the use of the term area factor. In addition, lessons learned from training MARSSIM show that describing the EMC concept in descriptive language, rather than by defining additional terminology, seems to improve understandability of the concept.

RESPONSE: I still see "area factor" referred to on Page 5-49, Line 8, and included in Example 14 in Section 8.63. The discussion in Appendix O on area factors is unnecessarily confusing. Recognizing the material in Appendix O is presented as "historical examples," the inclusion of the illustrative examples doesn't seem necessary, unless one were to include examples of how published area factors were misused in the past; such a discussion would make it much clearer why the use of "area factors" have been eliminated.

Frankly, the concept of residual material at the $DCGL_{EMC}$ over a small area being equivalent to residual material at the $DCGL_{W}$ over a large area is technically sound. I do not think Section 8.6.1 makes this point clearly. It is made on Page 8-21, Lines 1-6, and on Page 2-3, Lines 18-23, and Section 8.6.1 should reiterate some of that language.

I also note that in Chapter 4, on Pages 4-11 and 4-18, the text makes the point that the $DCGL_{EMC}$ is beyond the scope of MARSSIM, and strictly determined by regulatory agencies. I think this point should also be reiterated in Section 8.6.1.

Overall, the fractured approach to the $DCGL_{EMC}$ – i.e., it is outside the scope of MARSSIM, but referred to throughout MARSSIM, also makes the discussion in Appendix O on "area factors" superfluous, in my opinion, although as noted above, if it were revised to be a discussion on how *not* to use "area factors" in the context of a $DCGL_{EMC}$, it might have value, if and only if the other two points are also made again: 1) establishing a $DCGL_{EMC}$ is outside the scope of MARSSIM, and 2) why a $DCGL_{EMC}$ is useful, and should be discussed with the regulator.

3.3 Please comment on the effectiveness of the new organization of Chapter 4 (Considerations for Planning Surveys) to improve the understandability of the Chapter.

Earlier reviews of Chapter 4 provided evidence that the fundamental organization of Chapter 4 made it difficult to find and understand vital information. After discussing the challenge with experts in training and explaining the material, Chapter 4 was completely rewritten or reorganized in an attempt to improve understandability without changing the fundamental purpose of or material in the Chapter. In an effort to streamline the presentation of material in Chapter 4, some information was moved to Appendix O.

RESPONSE: I think Chapter 4 effectively presents the information that should be considered prior to undertaking the actual planning and design of surveys, as outlined in Chapter 5, with some exceptions noted in my second paragraph. I support the inclusion of the materials on the WRS and Sign Tests in Appendix O, but find the material on Area Factors unnecessarily confusing. Also, Sections O.5 and O.6 are just regurgitations of 4.12.8 and 4.12.9, respectively, so probably are not necessary.

It might be appropriate to remove the whole section on examples (Section 4.12) to Appendix O. I'd also consider moving the calculations of Section 4.5 to Appendix O; i.e., I'd keep some of the narrative discussion of the various DCGLs (a summary of each) in Section 4.5.3, but provide the details and calculations in Appendix O.

EDITORIAL NOTES:

Page 4-2, Line 31: This line notes DQO's were discussed in Chapters 1 and 2 and Appendix D, but should also note the discussion in Chapter 3 (Section 3.2).
Page 4-10, Line 14: "of all" should be "all of".
Page 4-11, Line 7: "…are beyond of the scope of MARSSIM." Delete "of".

Dr. Kenneth G.W. Inn

Question 1.1

1) Are the revisions to MARSSIM concepts and methodologies technically appropriate, useful and clear, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

1.1 The MARSSIM Workgroup wrote MARSSIM, Revision 1, for 1995 technology, not envisioning that future instrumentation would be able to measure a statistically significant portion of the survey unit while meeting required Measurement Quality Objectives (MQOs), especially that the Minimum Detectable Concentration (MDC)/Minimum Detectable Activity (MDA) be less than 50% of the Derived Concentration Guidelines Level for wide areas (DCGLw). New methods for designing, implementing and assessing scan-only surveys are included in the revisions to make effective use of resources when employing these technologies.

Earlier reviewers misinterpreted the term “scan-only surveys” to mean that samples would not be taken as any part of the survey process. Revision 2 has been further revised to indicate that quality control samples may need to be collected as part of the method validation or verification process, as appropriate.

$ *Please identify whether the inclusion and proposed implementation of scan-only surveys (Section 5.3.6.1 and Section 8.5) is appropriate, adequate and clear, especially the discussion on sampling for scan-only measurement method validation or verification.*

Response

Scan-only surveys adds an important tool for the implementation of MARSSIM, thus its inclusion for the proposed implementation is appropriate. It is always difficult to think that a survey measurement, and especially scan-only instruments, would be adequately quantitative to provide data for decision making. However, when the instrument offers adequate measurement uncertainty to meet the project quality MQOs, they become a viable tool. MARSSIM goes a long way to provide guidance for “what to do” but could do

better for "how to do it" with detailed examples and case studies. Please find here comments and recommendations that could improve technical appropriateness and usability of the document.

Section 5.3.6.1, p 5-42, lines 20-21 states "Scan-only surveys will likely require site-specific validation samples to ensure that the method can reliably detect concentrations at the DCGLW under the conditions expected at the site." This is a needed cautionary statement to tie the use of scan-only survey measurement capabilities back to the gray region and MQOs for the cleanup project.

> **Recommendation: MQOs need to be specified with the Calibration Source Provider to assure the appropriately commutable site-specific certified reference validation standards are prepared for accurate calibrations of direct measurement systems within necessary combined standard uncertainty limits that encompass the Type B uncertainty components particular to the scan-only measurements.**

5.3.6.1, p. 43, lines 7-17

Collecting samples for laboratory analysis to validate the scan-only measurements is a good QC practice.

> **Recommendation: The MQO for the laboratory analysis should have measurement uncertainty requirements that are at least 1/3 those for the scan-only measurements.**

5.3.6.1, p. 43, lines 16-24

The recommendations for good metrological practices enumerated here will help to support the validity of the scan-only option.

Additional Comment

Section 2.3.1, p. 2.12, lines 32-34

> **Recommendation: The DQO requirements for scan-only surveys should also call out measurement sensitivity requirements for scan-only MQOs.**

Section 8.5, p. 8-45, lines 1-18

> **Recommendation: The instructions for using Chelbyshev's inequality to determine**

a maximum UCL would benefit greatly by including examples to show how to use α and β with equation 8-3.

Section 8.5, p. 8-45 lines 19-24

"Chebyshev's inequality must be used with caution when there are very few points in the data set. This is because the population mean and standard deviation in the Chebyshev formula are being estimated by the sample mean and sample standard deviation. In a small data set from a highly skewed distribution, the sample mean and sample standard deviation may be underestimated if the high concentration but low probability portion of the distribution is not captured in the sample data set."

This is a great cautionary note.

Recommendation: Similar notes should be added to all statistical tools used by MARSSIM.

Question 1.2

1) Are the revisions to MARSSIM concepts and methodologies technically appropriate, useful and clear, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

1.2 Under hypothesis testing in MARSSIM, Scenario B is defined as assuming that the survey unit meets the release criteria unless proven otherwise, and its use was discouraged in MARSSIM, Revision 1. However, this is the only viable option for sites where the criterion is effectively “no added radioactivity” or “indistinguishable from background”. In Scenario B, the Lower Bound of the Grey Region (LBGR) is often set to zero, but the document allows use of a non-zero LBGR that considers background variability in determining whether the survey unit is indistinguishable from background. Since Scenario B assumes that the site meets the release criteria, there is a risk that the survey unit will pass simply because the survey did not have sufficient rigor. To guard against that, the revisions require that when using Scenario B, the survey unit must perform a retrospective power analysis to prove the survey has sufficient statistical power to detect a survey unit that should not have passed.

The non-parametric tests included in MARSSIM test the median instead of the mean. The release criteria are typically expressed as the mean. To guard against Scenario B situations where the median will pass but the mean won’t (this can occur in sample data distributions with a long tail in the higher concentration range), Revision 2 also requires that when using Scenario B, the survey unit must pass a quantile test to guard against excessive skewness.

Please comment on the inclusion and proposed implementation of Scenario B (Chapter 4, Section 5.3, and Chapter 8).

Question 1.2.a *Is it appropriate to recommend that Scenario B be used only for those situations where Scenario A is not feasible?*

Response

MARSSIM-2 is clear that evoking Scenario B is not to be taken lightly [e.g., 2.5.1, p. 2-29, line 23-41 and p. D-16, line 2-26]. As advanced measurement technology approach background levels, determination of essentially no residual radioactivity remaining in cleaned up sites become an exercise in distinguishing a value from "zero." Under these conditions Scenario A is not feasible, and Scenario B becomes the recommended option. To safeguard the Scenario B decision making process, multiple statistical tools are enlisted to provide the robustness necessary for scientifically defensible release of remediated sites. Make no mistake, this is a buyer beware situation where all affected parties will need to be fully knowledgeable of the application and limitations, and assumptions upon which the statistical tools are derived as a common basis for asking the cogent questions and making rational site release decisions.

Question 1.2.b *Are methods for considering background variability in assessing whether the site is indistinguishable from background reasonable and technically accurate?*

Response

The methods found that deals with background variability under Scenario B are the Kruskal-Wallis Test and the F-Test, which are technically accurate. Comparing the results from these two tests provide robustness to the decision about releasing the remediated site under Scenario B. However, the clarity of presentation is challenging. For example, Example 9 is built on equations and Tables from NUREG-1505, e.g. equations 13-3 and 13-13, and Tables 13.1 and 13.5.

Recommendations

The MARSSIM-2 Chapter 8 Example 9 should be stand alone so all steps can be clearly laid out and instructive for the reader, including the construction of Tables 8.12 and 8.13.

Additionally, the F-Test should also be worked out in the example.

Comment

In Example 9, $\sqrt{55} \times 3 = 0.74 \times 3 = 2.22$ for the LBGR])" should be "$\sqrt{0.55}$"

Question 1.2.c *Is the inclusion and proposed implementation of added requirements for retrospective power analysis and the Quantile Test while using Scenario B technically appropriate and discussed adequately and clearly?*

Response

The retrospective power analysis and Quantile Test are necessary to provide sufficient statistical robustness to confirm decisions of accepting or rejecting the null hypothesis for Scenario B, particularly for measurement data that are non-normally distributed. These statistical tests are discussed considerably, with retrospective power analysis referred to 33 times, and the Quantile Test is referred to 100 times. However, it is largely in statistician-ese and not so easily understandable. A lay-person understanding to the statistical tools would be very important to implement MARSSIM because so much of the decision making process is built on the statistical tools and their underlying assumptions.

Recommendation

Edit the manuscript of the statistician-ese, review the examples with lay-person step-by-step instructions so the reader can have full understanding of how to implement MARSSIM-2.

Question 1.3

1) Are the revisions to MARSSIM concepts and methodologies technically appropriate, useful and clear, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

1.3 Is the proposed implementation of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)? Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives? Is the method appropriate and practical for both laboratory and field (including scan) measurements? Please comment on the concerns of stakeholders that calculating measurement uncertainty for field measurements makes the survey process difficult to implement. In addition, please comment on whether recommendations provided by NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).

Question 1.3.a *Is the proposed implementation of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)?*

Response

MARSSIM-2 is a valuable resource for planning and executing site cleanups. There are over 300 references to "uncertainty" in the MARSSIM chapters and Appendices, many of which refer to variables affecting the uncertainty in the population distribution being assessed in the final site survey. However the novice is faced with understanding the right questions to ask, and figure out how to not be overcome by all of this information. MARSSIM is very good at providing guidance for "what to do," but could do better for "how to do." Clear and simple instructions with less statistical jargon, and detailed historical case study examples on how to make technical decisions would be very helpful for the reader - perhaps as an appendix chapter. Please find here additional comments and recommendations that could improve technical appropriateness and usability of the document.

Chapter 4 and Appendix D provide a rich discussion of factors that should be considered when developing MQOs. Many of these factors [e.g., reliability, robustness, dynamic range, sensitivity, specificity, calibration, specificity, sample collection, heterogeneity, attenuation, temporal effects, operator skill, speed, distance, blank, shielding] should be considers as uncertainty component inputs that would result in a comprehensive evaluation of project uncertainties, σ [total standard deviation of the population distribution being sampled] and u_{MR} [required measurement uncertainty]. Both of these uncertainties play a major role in determining the Δ/σ value used for the final site survey statistical assessments, and σ/u_{MR} used to set up the MQO for measurement capability needed. However, there are over 300 references to "uncertainty" in the MARSSIM chapters and Appendices, many of which refer to variables affecting the uncertainty in the population distribution being assessed in the final site survey. It is not difficult to imagine that information overload could be bewildering for the reader needing to determine σ and u_{MR} for the final site survey.

Recommendation: MARSSIM-2 should include detailed step-by-step worked out examples of setting up σ and u_{MR} **uncertainty component lists for a few cleanup scenarios with the expected measurement systems to be used. The examples should have topics that include:**

- **Measurement equation model for the selected survey scenarios and measurement processes**
- **Lists of Random [Type A] and Other [Type B] uncertainty components, expected values, sensitivity coefficients [partial differential calculus] and covariance effects [a tricky topic]**
- **Combining the uncertainty components into** σ and u_{MR}
- **Documenting the process of determining u, with acknowledgement to underlying assumptions**

- **Evaluating the σ/u_{MR} MQO**

The examples will need to be presented in plain-speak - that is devoid of statistical-ese and metrology-ese. With these examples, the reader will be better able to continue knowledgeable development of cleanup MQOs.

> **Recommendation: Stronger statements are placed in the MARSSIM text to alert the reader that factors discussed should be included in measurement uncertainty treatments.**

4.8.4.1, p 4-31, lines 1-3

"The required measurement method uncertainty is perhaps the most important MQO to be established during the planning process."

> **Recommendation: Great statement! However, the standard deviation of the population distribution being sampled is another very important MQO to be established during the planning process.**

Symbols, Nomenclature, and Notations, p. xxxii to xxxiv

The families of symbols for σ and u are extensive. Throughout MARSSIM, the σ symbols get a lot of use, but the u symbols do not. Furthermore, σ [theoretical total standard deviation of the population distribution being sampled] does not have an equivalent u symbol.

> **Recommendation: The u symbol should be included in the Symbols, Nomenclature, and Notations list. The u and σ symbols should be used where appropriate, particularly where Δ/u should be used.**

p. D-53, line 1-3

The uncertainty of a measurement expressed as combined standard uncertainty includes the counting uncertainty of the measurement instrumentation and the <u>sum of the errors</u> associated with the measurement system.

> **Recommendation: Edit the wording of the sentence. Equation 6-18 indicates that the <u>uncertainties</u> are combined as the <u>root-sum-of-squares</u>.**

Question 1.3.b *Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives?*

<u>Response</u>

The MARSSIM-2 use of measurement uncertainty is consistent with the concept of MQOs. It provides the quantitative basis for a risk minimization based site cleanup and verification process.

MARSSIM offers many tools for the determination of measurement uncertainties that supports the development and implementation of cleanup MQOs. However, the presentation of inputs and processes of calculating a measurement uncertainty is not cohesive which keeps the linkage of MQO and measurement uncertainties from being easily understandable.

For the experienced MQO and measurement uncertainty practitioner, MARSSIM-2 is a valuable resource for planning and executing site cleanups. There are over 300 references to "uncertainty" in the MARSSIM Chapters and Appendices, many of which refer to components affecting the uncertainty in the population distribution being assessed in the final site survey. However the novice is faced with understanding the right questions to ask, and figure out how to not be overcome by all of this information. MARSSIM-2 is very good at providing guidance for "what to do," but could do better for "how to do." Clear and simple instructions with less statistical jargon, and detailed historical case study examples on how to make technical decisions would be very helpful for the reader - perhaps as an appendix chapter. Please find here additional comments and recommendations that could improve technical appropriateness and usability of the document.

> **Recommendation: MARSSIM-2 includes detailed step-by-step worked out examples of estimating measurement uncertainty for a few measurement systems. The example should have topics that include:**

- **Project Gray Zone**
- **Project DQO**
- **Deciding on project σ and σ_{MR}**
- **Measurement MQO**
- **Selection of measurement instruments that could meet the project sensitivity requirements**
- **Measurement equation model for the selected measurement process and instruments**
- **Lists of Random [Type A] and Other [Type B] component budgets, expected values, sensitivity coefficients [partial differential calculus] and covariance effects [a tricky topic]**
- **Combining the uncertainty components into a standard combined uncertainty, u**
- **Documenting the process of determining u, with acknowledgement to underlying assumptions**
- **Evaluating the MQO that the measurement $u < \sigma_{MR}$**
- **If not, which uncertainty components can be improved for u to meet the requirements**
- **Determine $u_{NET\ BACKGROUND}$, critical level, and if necessary the MDC**

p. 6-31, Example 8

This example illustrates some of the steps used to estimate an uncertainty, in this case σ_y, for a measurement counting process. However, by adding the steps recommended above, the example

would become much richer and more illustrative to remove a lot of the mystery from the process of estimating measurement uncertainties, and also for estimating what's called the theoretical total standard deviation of the population distribution being sampled, σ, that is used for MQO Δ/σ.

These examples will need to be presented in plane speak -devoid of statistician-ese and metrologist-ese. With this example, the reader will be better able to continue with the process laid out in MARSSIM.

Question 1.3.c *Is the method appropriate and practical for both laboratory and field (including scan) measurements?*

Response

The use of measurement uncertainties is appropriate for the development of MQOs for both MARSSIM-2 laboratory and field measurements. For the experienced MQO and measurement uncertainty practitioner, MARSSIM-2 is a valuable resource for planning and execution of site cleanups. However the novice is faced with understanding the right questions to ask, and figure out how to not be overcome by all of this information. MARSSIM is very good at providing guidance for "what to do," but could do better for "how to do." Clear and simple instructions with less statistical jargon, and detailed historical case study examples on how to make technical decisions would be very helpful for the reader - perhaps as an appendix chapter. Please find here additional comments and recommendations that could improve technical appropriateness and usability of the document.

Recommendation: MARSSIM-2 should include detailed step-by-step worked out examples of estimating measurement uncertainty for a few measurement systems. The example should have topics that include:

- **Project Gray Zone**
- **Project DQO**
- **Deciding on project σ and σ_{MR}**
- **Measurement MQO**
- **Selection of measurement instruments that could meet the project sensitivity requirements**
- **Measurement equation model for the selected measurement process and instruments**
- **Lists of Random [Type A] and Other [Type B] component budgets, expected values, sensitivity coefficients [partial differential calculus] and covariance effects [a tricky topic]**
- **Combining the uncertainty components into a standard combined uncertainty, u**
- **Documenting the process of determining u, with acknowledgement to underlying assumptions**
- **Evaluating the MQO that the measurement $u < \sigma_{MR}$**
- **If not, which uncertainty components can be improved for u to meet the requirements**
- **Determine $u_{NET\ BACKGROUND}$, critical level, and if necessary the MDC**

p. 6-31, Example 8 This example illustrates some of the steps used to estimate an uncertainty, in this case σ_y, for a measurement counting process. However, by adding the steps recommended above, the example would become much richer and more illustrative to remove a lot of the mystery from the process of estimating measurement uncertainties, and also for estimating what's called the theoretical total standard deviation of the population distribution being sampled, σ, that is used for MQO Δ/σ.

These examples will need to be presented in plain speak -devoid of statistician-ese and metrologist-ese. With these examples, the reader will be better able to continue with the process laid out in MARSSIM.

Question 1.3.d *Please comment on the concerns of stakeholders that calculating measurement uncertainty for field measurements makes the survey process difficult to implement.*

Response

NIST and MARLAP relates measurement quantifiability to measurement uncertainty. Yes, guides like NIST Technical Note 1297 Guidelines for Evaluating and Expressing the Uncertainty of Measurement Results offers a methodology to estimate measurement uncertainty but it is not easy reading for the lay person. MARLAP's treatment of measurement uncertainty is consistent with that of NIST's but equally slow reading. MARSSIM offers many tools that could be used for the determination of measurement uncertainty, but the information is not presented in a cohesive process that makes the subject more understandable. It appears that MARSSIM makes the assumption that the reader has a working knowledge level of the subject, and uses referenced citations as its means of dealing with the nuances of uncertainty. However the novice is faced with understanding the right questions to ask, and figure out how to not be overcome by all of this information. MARSSIM is very good at providing guidance for "what to do," but could do better for "how to do." Clear and simple instructions with less statistical jargon, and detailed historical case study examples on how to make technical decisions would be very helpful for the reader - perhaps as an appendix chapter. Please find here additional comments and recommendations that could improve technical appropriateness and usability of the document.

Recommendations:

A. MARSSIM should use vocabulary consistent with NIST 1297 and ISO GUM, e.g.,

- **two types of uncertainty evaluations, i) Type A [statistical], and ii) Type B [other]**
- **the term "systematic uncertainty" should not be used, but rather, "Type B uncertainty".**
- **uncertainty components, u_i, contribute their "standard uncertainty" values to the**

uncertainty of the measurement result

- **the term "combined standard uncertainty", u_c, is used instead of "total uncertainty," σ_Y**

B. MARSSIM-2 should include detailed step-by-step worked out examples of estimating measurement uncertainty for a few measurement systems. The example should have topics that include:

- **Project Gray Zone**
- **Project DQO**
- **Deciding on project σ and σ_{MR}**
- **Measurement MQO**
- **Selection of measurement instruments that could meet the project sensitivity requirements**
- **Measurement equation model for the selected measurement process and instruments**
- **Lists of Random [Type A] and Other [Type B] component budgets, expected values, sensitivity coefficients [partial differential calculus] and covariance effects [a tricky topic]**
- **Combining the uncertainty components into a standard combined uncertainty, u**
- **Documenting the process of determining u, with acknowledgement to underlying assumptions**
- **Evaluating the MQO that the measurement $u < \sigma_{MR}$**
- **If not, which uncertainty components can be improved for u to meet the requirements**
- **Determine $u_{NET\ BACKGROUND}$, critical level, and if necessary the MDC**

p. 6-31, Example 8

This example is illustrative of some of the steps used to estimate an uncertainty, in this case σ_y, for a measurement counting process. However, by adding the steps recommended above, the example would become much richer and more instructive to remove a lot of the mystery from the process of estimating measurement uncertainties, and also for estimating what's called the theoretical total standard deviation of the population distribution being sampled, σ that is used for MQO Δ/σ.

These examples will need to be presented in plain speak -devoid of statistician-ese and metrologist-ese. With these examples, the reader will be better able to continue with the process laid out in MARSSIM-2.

C. For example for equation 6-5, σ_{bkg} in L_D should be expanded to include the uncertainty contribution from C.

D. *a postiori* MARSSIM historical records should be used to validate the MDC values used to see how applicable and accurate the concept was.

MARLAP recommendation of using the decision level [critical value = MARSSIM's critical

level] or confidence interval for the measurement result instead of the MDC/MDA is due to a large confusing and conflicting body of reports trying to apply the *a priori* MDC/MDA theoretical concepts that poorly documented the underlying assumptions and process. It is cleaner and simpler to base detection radionuclide content on combined standard uncertainties/expanded uncertainties at a stated confidence interval, or critical level for the determination of the presence of detectable radionuclide above background.

Question 1.3.5e *In addition, please comment on whether recommendations provided by NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).*

Response

Measurement quantifiability recommendations from NIST, ANSI/IEEE and MARLAP should be further incorporated into MARSSIM-2 rather than the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level. However, the MARSSIM-2 MDC/MDA is based only on the uncertainty of the NET background signal. This approach does not take into account all of the other sources of uncertainty affecting the measurements.

Recommendation: A sounder metrological approach would be to use the uncertainty of the population distribution being sampled [composed of measurement uncertainty components and sampling population].

Question 2.1

2. Does MARSSIM, Revision 2 provide useful, appropriate and clear examples and descriptions of technical approaches to implementing surveys and the statistics by which they are interpreted?

2.1

$ *Please comment on whether the description of updated measurement methods and instrumentation information (Chapter 6 and Appendix H) is useful, appropriate and clear.*

Response

The description of measurement methods and instrumentation information in Chapter 6 and Appendix H are generally useful and in large part appropriate and clear. However, there are descriptions of concepts and operations that could be made clearer and more useful. Please find here additional comments and recommendations in bold that could improve the technical usefulness, appropriateness and usability of MARSSIM-2.

6.2.2.3 Spikes and Standards

Recommendation: The spikes and standards should be confirmed for commutability.

6.3.1 Detection Capability for Direct Measurements, p. 6-9, line 2

[critical level, L_C; detection limit, L_D; minimum detectable concentration, MDC]

Comment

In the background counting space:

L_C = 2.33 * sqrt[B]; using NET background counts **(6-3)**

L_D = 3 + 4.65 * sqrt[B] = 3 + 2 * L_C; using NET background counts **(6-4)**

Equations 6-3 and 6-4 depend on the critical notion that B should be the mean NET background

counts, not the mean GROSS background counts. There are a number of instances in MARSSIM where B is used as the GROSS background counts and would mislead the reader, e.g., for estimating L_C, L_D, MDC, and could mislead defining DQOs and MQOs.

Recommendation: Carefully review MARSSIM-2 for the correct use of NET background counts, particularly for estimating L_C, L_D, and MDC. Furthermore, review the practice of not subtracting backgrounds and blanks counts from raw measured counts before calculating activity, Bq.

Comment

In the radionuclide concentration space:

MDC = C * L_D; and

(6-5)

C = conversion of cts to activity concentration

Recommendation

σ_{bkg} in L_D should be expanded to include the uncertainty in C

Comment: Example 1: [1/15 cm^2] should be [1 ct/15 cm^2]

6.4 Measurement Uncertainty

Recommendation

Make MARSSIM-2 vocabulary consistent with NIST/ISO GUM, e.g., Standard combined uncertainty, bias, systematic, Type 1 and Type 2 evaluations of uncertainty components

6.4.1 Systematic and Random Uncertainties, p. 6-28, lines 22-27

It is difficult—and sometimes impossible—to evaluate the systematic uncertainty for a measurement process, but bounds should always be estimated and made small compared to the random uncertainty, if possible. If no other information on systematic uncertainty is available, Currie (NRC 1984) recommends using 16 percent as an estimate for systematic uncertainties (1 percent for blanks, 5 percent for baseline, and 10 percent for calibration factors).

This seems to be a convenient solution for declaring values for systematic uncertainty components, but added effort needs to be invested to characterize them for a realistic estimation of the standard combined uncertainty and its documentation.

Recommendation: Consult with the metrologists and subject-matter experts for better estimates of Type B uncertainty components.

6.4.3 Uncertainty Propagation, p. 6-29, lines 8-12

Assuming the individual uncertainties are relatively **small**, symmetric about zero, and independent of one another, then the total uncertainty for the final calculated result can be determined by solving the following partial differential equation: equation 6-18

Comment: When uncertainties are on the order of 10-20%, the second order effects to the Taylor Expansion are needed. What happens to equation 6-18 is likely to be even messier when the uncertainties get even larger when approaching the MDC. Utilizing measurement technology with higher sensitivity may offer a viable option to tame expansion of equation 6-18.

6.4.4 Reporting Confidence Intervals, p. 6-31, line 1

Example 8 illustrates the point for computing an expanded uncertainty at a given confidence level. However, the number of background counts is only one part of a measurement equation and/or a whole uncertainty menu.

Recommendation: A more complete example would be much more instructive, taking the reader through the whole process: measurement equation, additional sources of Type B uncertainties, uncertainty component menu, estimating values for uncertainty components, sensitivity factors, correlated uncertainties, combining the uncertainty components into the standard combined uncertainty, expanding the uncertainty at confidence levels, and reporting the uncertainties.

6.5 Select a Service Provider to Perform Field Data Collection Activities

MARSSIM recommends that this selection take place early in the planning process so that the service provider can provide information during survey planning and participate in the design of the survey. Service providers may include in-house experts in field measurements and sample collection, health physics companies, or environmental engineering firms, among others.

Comment: Include radioanalytical service representatives as well.

MARLAP Chapter 5 Lab Services also provides guidance for selecting a service provider, including Pre-award proficiency evaluation, assessment and audit.

An appropriate accreditation program, based on MARSSIM-2, could provide consistency among capabilities, operations, quality of results that could relieve programs from

conducting these quality improvement tasks on a case by case basis.

6.6.1.2 Direct Measurements [fixed measurements in the field], p. 6-34, lines 25-28

Determining a calibration function and the associated MDA/MDC can be complicated and may contribute to the total measurement method uncertainty, especially for situations other than radionuclides uniformly distributed on a plane or through a regularly shaped volume (e.g., a disk or cylinder).

The metrologists and subject-matter experts have tools that can help deal with these complicated scenarios.

Recommendation: Consult with MARLAP secondary reference laboratory and source producers to establish DQO/MQO for calibration sources

6.6.4 Instrument Calibration, p. 6-38, line 32

If the actual field conditions differ significantly from the calibration assumptions, a special calibration for specific field conditions may be required. Such an extensive calibration need only be done once to determine the effects of the range of field conditions that may be encountered at the site.

Comment: When field conditions change, recalibration should be done for each measurement system used using commutable certified reference sources that accommodate the new measurement conditions. Alternatively, when possible, computational bias corrections and uncertainties estimates could be applied to the measurement systems to accommodate the new effects on the measurement systems. Furthermore, calibrations should be checked at a prescribed interval, and QC check sources used to assure continued and stable operability between calibrations to continue to comply with the MQO requirements.

6.6.4, p. 6-39, lines 14-24

Calibration sources should be traceable to NIST. Where NIST-traceable standards are not available, standards obtained from an industry-recognized organization (e.g., the New Laboratory for various uranium, thorium, and plutonium standards) may be used. Calibration of instruments for measurement of residual radioactive material on surfaces should be performed such that a direct instrument response can be accurately converted to the 4π (total) emission rate from the source. An accurate determination of activity from a measurement of count rate above a surface in most cases is an **extremely complex task** because of the need to determine appropriate characteristics of the source, including decay scheme, geometry, energy, scatter, and self-absorption. Proper calibration ensures that systematic errors in measurements are controlled to help ensure that the MQO for measurement method uncertainty is met.

Recommendation: MQOs need to be specified with the Calibration Source Provider to assure the appropriately commutable certified reference standards are prepared for accurate calibrations of direct measurement systems within necessary combined standard uncertainty limits.

6.6.4, p. 6-40, lines 35-40

Source efficiencies may be determined experimentally. Alternatively, ISO-7503-1 (ISO 1988) makes recommendations for default source efficiencies. A source efficiency of 0.5 is recommended for beta emitters with maximum energies above 0.4 megaelectronvolts (MeV). Alpha emitters and beta emitters with maximum beta energies between 0.15 and 0.4 MeV have a recommended source efficiency of 0.25. Source efficiencies for some common surface materials and overlaying material are provided in NUREG-1507 (NRC 1997a).

Recommendation: This section refers to ISO-7503-1 (ISO 1988) recommendations for source efficiencies. The values should not be used without uncertainty estimates associated with them. These uncertainties would be considered Type B that need to be folded into the determination of standard combined uncertainties.

Monte Carlo simulations that takes into account source related factors such as type of radiation and its energy, source uniformity, surface roughness and coverings, and surface composition (e.g., wood, metal, concrete) could be helpful in estimating the source efficiency. Regardless, estimating a source efficiency without actual measurements must be accompanied with an uncertainty estimate that is to be incorporated into the standard combined uncertainty of the measurements made by the measurement systems. As a minimum a rectangular uncertainty distribution could be used.

6.7.1, p. 6-47, Table 6.7

Recommendation

See changes in <u>bold</u>

Radiation Type	Hand-Held Instruments	In-Situ Gamma Spectroscopy
Direct Measurements		
Alpha	**<u>Poor</u>**	NA
Beta	**<u>Fair</u>**	NA
Photon	**<u>Fair</u>**	**<u>Fair</u>**
Neutron	Good	NA

Scanning Survey		
Alpha	**Very Poor**	NA
Beta	**Poor**	NA
Photon	**Fair**	**Fair**
Neutron	Fair	NA

6.7.1, p. 50, line 4: Please change "Bq/m^2" to "cps/m^2"

6.7.1, p 6-51 line 4: "Equation 6-17" should be "Equation 6-19"

6.8, p. 6-51, line 12

Example 9 is an opportunity to demystify handling the concept of uncertainties.

> **Recommendation: Include how the estimated uncertainty of the overall efficiency would be calculated**

6.7.2

> **Recommendations: Analytical procedures should also include major correction factors for blank, decay to a reference time, and radiation emission probability for the conversion of counts to activity units and application of MDC.**

> **Ra bearing minerals would also be a source for Rn, and should be included in the discussion.**

> **The discussion of estimating radiation dose from Rn exposure should be part of the determination of the AL and DCGL. Furthermore, the effects of Rn and progeny on fixed and survey measurements of other radionuclides should be discussed since they could be a source of measurement interference.**

6.8

Good discussions with recognition of many factors that complicates Rn measurements. These complicating factors should be recognized in the estimation of measurement uncertainty.

Global Recommendation: MARRSIM should collect all of the recommendations and put them in a list at the end of each Chapter

Comments on Appendix H

Appendix H needs to provide the pros/cons/applications information for the reader in a consistent way to both the text and tables. Sometimes the information is in the Description column, sometimes in the Applications column and sometimes in the Remarks column. Set up a consistent organization so the reader can find the information more systematically.

Some Pro/Con to add to the Tables may include attributes such as:

Table H.2

Are radiochemical separation or other preparation needed before counting? Will they need specialized laboratory, reagents, air exhaust, waste handling?

Is vacuum needed?

Is cooling needed, e.g., mechanical, LN2?

Is stable electrical source needed?

Is vibration stability needed?

Power requirements?

Space requirements?

Floor loading support needed?

Contamination prevention?

Cleanroom environment needed?

Shielding/attenuators needed?

Calibration requirements?

Background requirements?

Radiation interference?

Specialty gas?

Room air requirements?

Detector/sample geometry requirements/stability/reproducibility?

Detector jigs?

Specification for standardization? Commutability?

Performance specs, e.g., efficiency, energy range, energy linearity, resolution, S/N ratio, spectral deconvolution, dead time, specialized operator/radiochemistry training, etc

Table H.3

Weight?

Size?

Ruggedness requirements, e.g., temperature, humidity, electrical interference, pressure, etc?

Contamination prevention?

Shielding/attenuators needed?

Calibration requirements?

Detector/sample geometry requirements/stability/reproducibility?

Detector jigs?

Specification for standardization? Commutability?

Performance specs, e.g., efficiency, energy range, energy linearity, resolution, S/N ratio, dead time, specialized operator training, etc

Table H.4

Weight?

Size?

Is cooling needed, e.g., mechanical, LN2?

Contamination prevention?

Shielding/attenuators needed?

Calibration requirements?

Detector/sample geometry requirements/stability/reproducibility?

Detector jigs?

Specification for standardization? Commutability?

Performance specs, e.g., efficiency, energy range, energy linearity, resolution, S/N ratio, spectral deconvolution, dead time, specialized operator training, etc

Table H.5

Weight?

Size?

Ruggedness requirements, e.g., temperature, humidity, electrical interference, pressure, etc?

Contamination prevention?

Shielding/attenuators needed?

Calibration requirements?

Detector/sample geometry requirements/stability/reproducibility?

Detector jigs?

Specification for standardization? Commutability?

Performance specs, e.g., efficiency, energy range, energy linearity, resolution, S/N ratio, spectral deconvolution, dead time, specialized operator training, etc

Table H.6

Weight?

Size?

Ruggedness requirements, e.g., temperature, humidity, electrical interference, pressure, etc?

Contamination prevention?

Calibration requirements?

Detector/sample geometry requirements/stability/reproducibility?

Detector jigs?

Specification for standardization?

Performance specs, e.g., efficiency, energy range, energy linearity, S/N ratio, specialized operator training, etc

Table H.7

Are radiochemical separation or other preparation needed before counting? Will they need specialized laboratory, reagents, air exhaust, waste handling?

Is vacuum needed?

Is stable electrical source needed?

Is vibration stability needed?

Power requirements?

Space requirements?

Floor loading support needed?

Cleanroom environment needed?

Room air requirements?

Specification for standardization? Commutability?

Performance specs, e.g., transmission efficiency, isotopic fractionation effects, mass linearity, resolution, sensitivity, detector efficiency, etc

Some text information is not consistent with that presented in the Tables, e.g., information in Table H.1 is not consistent with that in Table H.7

Table H.8

Are radiochemical separation or other preparation needed before counting? Will they need specialized laboratory, reagents, air exhaust, waste handling?

Is vacuum needed?

Is cooling needed, e.g., mechanical, LN2?

Is stable electrical source needed?

Power requirements?

Space requirements?

Floor loading support needed?

Contamination prevention?

Shielding/attenuators needed?

Calibration requirements?

Background requirements?

Radiation interference?

Detector/sample geometry requirements/stability/reproducibility?

Detector jigs?

Specification for standardization?

Performance specs, e.g., efficiency, energy range, energy linearity, resolution, S/N ratio, spectral deconvolution, dead time, specialized operator/radiochemistry training, etc

Dr. Annie Kersting

Overall: I found the MARSSIM report useful and clear. It delivers on its promise to provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces. I found the report logical, well written and straightforward in its implantation approach.

1.4 Is the discussion of survey requirements for areas of elevated activity technically accurate, appropriate and clear? In particular, please comment on the decision to maintain the use of the unity rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4). Are there suggested alternatives to the use of the unity rule?

> **While modeling is outside the scope of MARSSIM, depending on the modeling tool or methodology used to develop release criteria, the use of the Unity Rule for multiple areas of elevated activity in a single survey unit can lead to unrealistic or overly conservative assumptions. For example, the models may assume that the receptor is located directly above each area of elevated activity and stays there for the duration of their exposure period. This physically cannot occur in cases where there is more than one area of elevated activity per survey unit and results in concerns that this will cause an over-estimate of dose or risk, leading to an emphasis on remediating areas of elevated activity that don't incur additional significant dose or risk to receptors.**
>
> **MARSSIM, Revision 2, does not change recommendations for the use of the unity rule, but emphasizes assessing whether criteria for areas of elevated activity apply to survey units, and when they do, using a common sense approach to applying these criteria, keeping in mind the limitations of the unity rule described above for multiple areas of elevated activity.**

Response:

The survey requirements covered in Sections 4 & 5 are logical, clearly written and the technical sections are appropriate and useful. Overall, Sections 4 & 5 provide a practical and comprehensive approach to completing an environmental radiological survey or surface soil and buildings.

Specific comments:

- Section 4.3.6 A note on subsurface assessment:
  - I would suggest repeating the definition of surface vs subsurface depth, to remind the reader.
- Section 4.4:
  - The utility rule is well described.
  - L5. The sentence: Essentially, this means that if measurements….helps clarify the utility rule.
  - L15-16: to slightly clarify this paragraph, I would state…..each fraction, (f), is determined…..

- Section 4.5.3.7: Small areas of Elevated activity
  - Although discussed in 5.3.5, (and 4.2.5 as well) elevated activity in multiple areas and the Utility Rule also discussed in this section and is appropriate. It references 5.3.5.1 and .2.
- Section 5.3.5 Establishing procedures for elevated activity discusses using the combined impact of each elevated area. This section logically discusses how to determine if sufficient data points have been taken, and Example 7 & 8 shows a step by step example
  - When evaluating areas of elevated activity, the text is clear about reminding the reader to keep in mind the limitations of the utility rule.
- Section 8: Interpretation of survey Results
  - 8.6: discusses the results, elevated measurements and statistical tests.
    - 8.6.3 discusses what to do if the survey unit fails and lays out possible options, with a given example (13). Example 13-15 are appropriately illustrative.
    - Example 15 was not as clear as the others. It uses the Wilcoxon Ran Sum test. Is this the only test available, and are the assumption that go into the WRS appropriate for all Class 1 failures?
  - Appendix 0..Wilcoxon Ran Sum test
    - It might be helpful to mention Appendix O in section 8.6 in example 15 above
  - Appendix O.4. My background in statistics is not sufficient to evaluate the details of this section and the examples presented.

1.5. Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles technically sound and appropriate, and is the description accurate? In particular, please comment on the rule-of-thumb for determining when use of MARSSIM may not be appropriate for survey units containing discrete radioactive particles (Section 4.12.8 and Appendix O.5).

Discrete radioactive particles have an extremely small size and contain enough activity that survey units containing discrete radioactive particles generate impractical survey designs under MARSSIM. Over MARSSIM's twenty-year history, several sites have attempted to utilize MARSSIM to address discrete radioactive particles, with predictably extreme survey designs as a result. In addition to being impractical, designs for discrete radioactive particles violate some of the assumptions commonly made during modeling, which includes parameters based on an areal source of radioactive material, e.g., length of the area of the elevated activity in the direction of overland flow. While modeling is outside of the scope of MARSSIM, it is nonetheless required that survey designs match the assumptions made during modeling, otherwise, the survey design does not meet the requirements of the action level.

To set a limit for determining when areas of elevated activity are too small to use the traditional MARSSIM methodology, the MARSSIM Workgroup used a traditional rule-of-thumb for instrumentation. When the length of the area of elevated activity is less

than three times the distance to the detector, the area of elevated activity is viewed by the detector as a point source instead of as an areal source. These point sources will need different receptor modeling and release requirements, and hence different survey designs than traditional areal sources.

At this time, MARSSIM does not provide guidance on designing discrete radioactive material surveys. It is the intention of the revision that additional information provided should prevent MARSSIM from being applied inappropriately to survey units involving discrete radioactive particles.

Response:

Hot particles or discrete radioactive particles are and important discussion that needs to be addressed separately from average survey sampling. Section 4.12.8 is mostly clear on the importance of this topic and the discussion is useful accurate.

- Section 4.12.8. Release Criteria of radioactive particles
  - A point of confusion in the Charge Question text above, I thought designing a survey for discrete material would be out of scope from MARSSIM.?? Has this section been applied inappropriately to survey units in the past, say by users from 2000-2010?
  - It would be beneficial to add text if we know where the users had concerns or confusion. Any additions might be best after a discussion by the committee.
  - Rule of thumb d> 3L (equation 4-26)
    - I agree that there needs to be some general guidance, but I don’t have any context as to why this particular equation was chosen, so can’t say whether or not the this is best equation to assess if the assumptions inherent in the dose or risk model are violated. I would think that the identity of the particle (alpha, gamma, etc) and its associated activity would have an impact in any useful equation.
    - L14-16. I find this vague, as no guidance is given as to determining the necessary concentration of discrete particles that are necessary. Maybe this paragraph should be expanded.
- Appendix O.5: release criteria for discrete radioactive particles
  - L4-10. This paragraph provides clear and relevant context for why hot particles exist and their importance. This section first discussed in 4.12.8 should be repeated again here.
  - L 11-18, the second paragraph is equally important for explaining why the MARSSIM EMC process is not a valid approach. This section is clear and relevant and also importantly repeated.
  - It appears the Appendix 0.5 is mainly a repeat of 4.12.8 and I’m not sure what new information is added or expanded on to provide more value.

Dr. Amy Kronenberg

Question 1.1

Is the inclusion and proposed implementation of scan-only surveys (section 5.3.6.1 and section 8.5) appropriate, adequate and clear, especially the discussion on sampling for scan-only measurement method validation and verification.

Response: Scan-only technology can be considered for certain types of sites and conditions, however, the depth of the description of these technologies is weak in comparison to the more traditional measurement approaches with well-vetted detector technologies and analytical methods that are extensively discussed in the draft Revision 2 of MARSSIM and its draft Appendices. A more robust description of the relevant equipment, including how the equipment maps to DQO and MQO's for a survey conducted with scan-only approaches should be added, and examples of read-outs of sample data sets from representative scan-only technology should be provided. While the properties necessary for the acceptance of a scan-only technology are described for different classes of survey areas, the overall discussion of this approach requires a more complete description to support the technology advances since the last revision of MARSSIM.

A more extensive description and example of method validation is also warranted, including number of replicate measurements over defined scanning areas and appropriate QA/QC definition for the scanning technology and its recording output.

Verification of recorded values using a second direct measurement approach is described and it is clear that this is expected as a supplement to scan-only surveys.

Question 2.1

Is the description of updated measurement methods and instrument methods and instrumentation information (Chapter 6, Appendix H) useful, appropriate and clear.

The content of Chapter 6 (Field Measurement Methods and Instrumentation) is laid out well with the inclusion of excellent tables, examples and sample calculations. The chapter presents common sense approaches for how to choose the best detection method for a given radionuclide that may emit more than one type of radiation, and an excellent discussion of measurement uncertainties and how to manage them. Only one minor comment – table 6. 8 considering advantages and disadvantages of instrumentation and measurement techniques should spell out the name of an ISGS (in situ gas spectrometer) to avoid having to look up the acronym.

Appendix H is written clearly and provides more detailed information to support instrumentation choice for particular needs.

Question 3.1

Please comment on the revised description of how to set the Lower Bound of the Gray Region (LBGR) and its likely effectiveness in encouraging users to rely on site-specific information for doing it (Chapter 4 and section 5.3).

For the naïve MARSSIM user, the concept of the LBGR remains challenging. It is introduced as a quantity in use in an example in Chapter 4 without being properly defined in advance. The central question that arises is what is the purpose of the LBGR? Why is it needed? Why in the examples provided is the LBGR "usually set" at the median concentration of the radionuclide? How can the median concentration of a radionuclide be known when a survey is first being designed without prior measurements and only limited historical information is available? This is not clear to a technically trained reader who is not deeply familiar with the MARSSIM process. Why in section 5.3.1 is there a statement that for scenario A, survey designers choose a discrimination limit (DL), the lower bound of the gray region (LBGR)? Are these equivalent? Again, what is the purpose of the LBGR? These statements appear prior to the section in which the Gray Region is illustrated schematically for Scenario A and Scenario B, but overall the document suffers from a lack of clarity on the purpose of the LBGR with regard to radiation protection and disposition of a site undergoing evaluation for release.

Dr. Robert Litman

Response to Charge Question 1.3

1.0 Are the revisions to MARSSIM concepts and methodologies **technically appropriate, useful and clear**, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

1.3 **RESPONSE**

The narrative for question 1.3 has several parts that, although related, are separate questions. In order to focus in on each aspect, the question has been subdivided into the following four parts.

A. Is the proposed implementation of the of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)?
B. Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives?
C. Is the method appropriate and practical for both laboratory and field (including scan) measurements? Please comment on the concerns of stakeholders that calculating measurement uncertainty for field measurements makes the survey process difficult to implement.
D. In addition, please comment on whether recommendations provided by NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).

A. Is the proposed implementation of the of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)?

No. The most important measurement quality objective must be <u>method uncertainty</u> based on the project DQOs.

- Section 4.8.2 contains references to other documents and a list of some MQOs. The first bullet on the list for measurement quality objectives, required method uncertainty, also needs the qualifiers that it is for a specific matrix. As an example, soil is not a specific matrix. It could be loam, sand, clay or gravel. If using scanning instruments, the average atomic number of the matrix will determine factors like self-shielding, which affects the measurement uncertainty and the detection efficiency. This is missing from the list and any discussion in MARSSIM. The remainder of any discussion of measurement uncertainty is deferred to the discussion in Section 6. In section 6.4 on page 6-27 line 4 the following is stated:

> "The FSS data, which is used to document the final radiological status of a site, *should* state the uncertainties associated with the measurements. Conversely, detailing the uncertainties associated with measurements made during scoping or characterization surveys *may or may not be of value*, depending on what the data will be used for, as defined by the DQOs." (I have added italics for emphasis).

- Any measurement, no matter what its intended use is, *shall have* an uncertainty associated with it, otherwise it is a useless number. I recognize that MARSSIM is a guidance document, however, guidance of this type Stating that uncertainty may or may not be of value, is not scientifically sound.
- Section 4.8.3.1 discusses scanning measurements and says,

> "Items that should be kept in mind while investigating a scan-only survey are—
>
> - data and location logging
> - reproducibility of the measurements (e.g., fixing a detector at a constant distance from a surface)
> - MDCs
> - scanning speed and operator training
> - data integrity and security
> - selecting an appropriately sized area for elevated measurement comparison calculations"

All of these 'items' affect the most basic MQO of measurement uncertainty. MARSSIM must recommend that if a scan-only survey is used, that these 'items' shall be included.

- Appendix D.1.5 on line 27 states,

> "Negative attributes include that (1) it is not a very representative measure of central tendency for highly skewed distributions, and (2) it is not useful when a large proportion of the measurements are reported as less than the detection limit (EPA 2006b)."

The meaning of part (2) of the statement is unclear. Any measurement of radioactivity shall be reported as calculated or read directly from an instrument <u>with</u> its associated uncertainty. At no time shall results be reported as"<", "MDC", "0", or "ND". These designations have no mathematical basis and cannot be accounted for in statistical evaluation of data (this is notionally stated on page D-52 of MARSSIM, "Results reported as "<MDC" cannot be fully used and, for example, complicate even such simple analyses as calculating an average"). Can a survey unit have values that are less than the detection limit? The answer to that is of course yes if there is no contamination. The reporting of calculated values whether negative or positive is in line with MARLAP and the GUM. Rather than

using the above statement for the user to possibly misconstrue as license to accept non-numerical values, suggest using the following wording:

> "Mean values belonging to distributions that are skewed, or bimodal in nature should not be used for data assessment or hypothesis testing. Rather a different type of statistical assessment or a subdivision of the area classification should be considered."

The next paragraph should be shortened to read as follows:

> "The median is a value that corresponds to the "center" of a distribution, but where the mean represents the center of gravity, the median represents the "middle" value of a distribution. The median is that value such that there are the same number of measurements greater than the median as less than the median. The positive attribute of the median for data assessment and hypothesis testing is that it is useful when the action level is based on long-term, mean health effects".

- The example described using Figure D.3 is confusing. The top figure has the ordinate labelled as "Distribution of Residual Radioactive Material" and the ordinate labelled as "Survey Unit" and also "Concentration" with a large space between the two abscissa descriptions. Generally, the distribution in a survey unit would refer to a surface area which should yield a two-dimensional ordinate. Is the ordinate in D.3 merely trying to show how many samples have a concentration above zero? It also appears that the figures are attempting to depict a Normal distribution. However, the figures have an odd inflection point location.

  The second part of Figure D.3 is even more confusing. Here again the labelling of the axes is confusing. Does

  > "f(δ) is the sampling of the estimated survey unit mean"

  Refer to the ordinate values being the distribution of the actual sample values?

  The narrative that follows the figure discusses f(δ) on the ordinate being compared to the DGCL value on the abscissa. It assumes that the distribution in the survey unit follows the Normal distribution pictured. Accepting that is true, why not have a simple statement that says, "It is clear from the graph for this hypothetical case that the survey unit mean is less than the $DCGL_W$ and thus the survey unit is in compliance based on the decision rule."? Or is the example attempting to show something else?

B. Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives?

No. Values for method uncertainty must be derived from Project DQOs. MARLAP specifically identifies how this is to be accomplished and then used to find acceptable methods of analysis. This is not used in MARSSIM

- The measurement method uncertainty is defined in section 2.3.1, on page 2-13. This definition relies on optimal conditions of sample measurement:

> "Reasonable values for measurement method uncertainty can be predicted for a particular measurement technique based on typical values for specific parameters (e.g., count time, efficiency) and previous surveys of the areas being investigated "

It is not clear from this definition how this uncertainty could relate to a project DQO for measurement uncertainty determined before sampling has been done. This method of uncertainty determination allows the user to select a measurement method after DQOs have been established. This definition is also not aligned with the definition of *required method uncertainty* as defined in MARLAP. If MARSSIM, MARSSAME and MARLAP are to be used as supporting documents for each other during site investigations, the definitions/calculations of method uncertainty should be the same.

C. Is the method appropriate and practical for both laboratory and field (including scan) measurements?

No.

- The most important MQO is that of measurement uncertainty. This MQO must be the result of the stakeholder's agreement on the defensibility of the DQOs based on the field and laboratory measurements. The fact that calculating measurement uncertainty is difficult is not a reason to not perform the calculation. A well-developed project quality assurance plan should identify the calculation process and also the quality control checks necessary to assure that the measured uncertainties will support the defense of the decision-making process.

D. In addition, please comment on whether recommendations provided by NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).

The current recommendations need to be changed to reflect the importance of selecting a method uncertainty first before making measurements. Values of MDC or MDA are not detection values. These are only best guesses at what potentially can be determined.

Only *a posteriori* determinations based on critical level and measurement uncertainty can determine detectability.

- The MDC/MDA *should not* be set at 10-50 % of the action level, and should not be used in conjunction with sample detection. MARLAP training for the past 15 years has emphasized the use of a discrimination level (DL) for the lower bound of the gray region. The DL has to be a value that is attainable with the detection method used and is a positive result for the radionuclide of interest. The MDC does not have any direct relation to the DL. There exist multiple equations for the determination of the MDC, and in almost every case a value that meets the MDC has a probability of being a true detection 50 % of the time. MDC is not used as an assessment of sample detection. MARLAP states the following:

  > "It is difficult to imagine a scenario in which any useful purpose is served by comparing a measured result to the MDC"

- On page D-54 of MARSSIM, the following is used to explain how to use data qualifiers for results that are less than the MDC:

  > "The following are examples of data qualifier codes or flags derived from national qualifiers assigned to results in the contract laboratory program: a normal, not detected (less than critical value) result (U) or <MDC. The sample was analyzed for the radionuclide of interest, but the radionuclide concentration was below the MDC. MARSSIM recommends reporting the actual result of the analysis, so this qualifier would inform the reader that the result reported is also below the MDC."

This uses the MDC as a means of qualifying detection which is a direct contradiction to the purpose of an MDC as an *a priori* assessment of measurement detection, not a determination of detection following sample measurement.

Response to Charge Question 1.4

1.4. **Is the discussion of survey requirements for areas of elevated activity technically accurate, appropriate and clear?**

Survey Requirements

RESPONSE

The discussion of survey requirements regarding determination of area factors and the number of samples to be collected is technically correct.

- However, the following discussion,

  "If the actual scan MDC is less than the required scan MDC, no additional sampling points are necessary for assessment of small areas of elevated activity. In other words, the scanning technique exhibits adequate detection capability to detect small areas of elevated activity."

  The conclusion here depends upon how the scan MDC is determined, i.e., what equation is used for that determination? This goes back to the fundamental issue of uncertainty in the scanning instruments measurement process and how that uncertainty is measured. If a large method uncertainty is acceptable then the actual measurement values with their associated uncertainty may exceed the $DCGL_{EMC}$. If too small a method uncertainty is required, an unusually long time for surveys or much more costly instrumentation may be required.

  How much less than the required scan MDC does the instrument MDC need to be so that it can be reasonably assured that the area(s) of elevated activity are properly accounted for in the larger scope of exceeding the $DCGL_W$.

Editorial Comments

- The Examples 7 and 8 provided in Section 5.3.5 demonstrate how to use the information discussed in 5.3.5. however, there are assumptions made for the initial calculation of the number of samples to be taken that should be stated in the example so that how the numbers were determined is clear to the reader.

Unity Rule

In particular, please comment on the decision to maintain the use of the unity rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4). Are there suggested alternatives to the use of the unity rule?

RESPONSE

The use of the unity rule should be maintained in the document. I don't believe that alternatives can exist to *accurately reflect* the potential for dose assessment.

- Section 8.6.2 discusses the unity rule equation to be used if more than one area of elevated activity exists

  > "If there is more than one elevated area, a separate term could be included in Equation 8-4 for each area. The use of the unity rule for more than one elevated area may imply that a person is centered on each area of elevated radioactive material and exposed simultaneously. This is an impossible situation and represents a very cautious exposure scenario."

  The way the unity rule is used in this discussion is for only one radionuclide is present, and the same one at each area of elevated activity. If that is true, then the unity rule as stated is certainly overprotective, but it is in agreement with other agencies rules of providing conservative dose assessment. However, for the case of multiple radionuclides that may be unrelated, or not in an equilibrium chain (where an exact determination of radionuclide ratios may possibly be determined), the unity rule would provide a more reasonable estimate of dose since each would have a different $DCGL_{EMC}$.

Overall, maintaining the unity rule in MARSSIM is a consistent and conservative approach to maintain reasonable dose assessment.

Response to SAB Charge Question 1.5

1.5. Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles technically accurate, appropriate and clear? In particular, please comment on the rule-of-thumb for determining when use of MARSSIM may not be appropriate for survey units containing discrete radioactive particles (Section 4.12.8 and Appendix O.5).

RESPONSE

No. A site contaminated with DRPs cannot conform the basic tenets proposed by MARSSIM. DRPs do not uniformly distribute and the radionuclides contained within a DRP cannot be relied upon to exhibit specific ratios that can be used for surveys with scanning instrumentation. Appendix section O.5 does not appear to be necessary.

- There exist a few different mechanisms for forming discrete radioactive particles (DRP). While an exact definition of particle size and dose rate of such particles does not exist, the potential presence or absence of DRPs and what constitutes such for the survey, should have been identified as part of the historical site assessment process and the DQO process. The discussion in Section 4.12.8 states:

  > "…it is not acceptable to use the MARSSIM EMC process when the distance to the detector is greater than three times the longest dimension of the area of elevated activity, as represented by Equation (4-26): $d>3L$"

  This particular statement and the equation may make the user assume that every time they come across a part of the survey unit with a DRP that it can be treated in this manner. What happens if the DRP areas overlap? How many are acceptable? The equation also assumes that a single radionuclide may be responsible for the elevated reading. Does this method consider the possibility of alpha, beta and gamma emissions of multiple radionuclides as part of the DRP?

- The bulleted items that start on line 12 of page 4-66 appear to be an *a posteriori* thought process rather than part of the original DQO process. Especially the statement,

  > "If discrete radioactive particles do not contribute significantly to dose or risk at a site, it is a reasonable assumption that they will not affect the outcome of a wide-area FSS."

  This can't possibly be known until the DRPs are located and assessed based on the DQOs.

If one DRP is present there certainly will be more. The real question is how to plan a survey where in advance you know that there will be multiple DRPs. The MARSSIM as currently written cannot accommodate that occurrence. This section and MARLAP in general, should state this in much stronger terms.

- Appendix O.5 merely repeats all the information already provided in Section 4.12.8. There does not seem to be any purpose to O.5.

Response to SAB Charge Question 2.1

2.1 Please comment on the usefulness and accuracy of updated measurement methods and instrumentation information (Chapter 6 and Appendix H).

RESPONSE

Chapter 6

Chapter 6 is useful but should be edited to include the items indicated below. It is also important the concept of MDA or MDC not be linked to detection in any manner. This must be made clearer throughout the document.

a) Chapter 6 describes the theoretical equations applied to the concept of the minimum detectable concentration (MDC) and how it is applied to field measurements using scanning or direct surface measurements. The reference data used for much of the background information for this chapter is provided in NUREG-1507. The following is an excerpt from that reference with regards to the establishment of controlled conditions for measuring the MDC:

> "A measurement hood, constructed of Plexiglas, provided a controlled environment in which to obtain measurements with minimal disturbances from ambient airflow. The Plexiglas measurement hood measured 93 cm in length, 60 cm in height, and 47 cm in depth, and was equipped with a barometer and thermometer to measure ambient pressure and temperature within the chamber. Measurements were performed within the measurement hood using a detector-source jig to ensure that the detector-to-source geometry was reproducible for all parameters studied."

It is important to note that the reference also states the following:

> "Therefore, the terms MDC and detection limit should not be used interchangeably."

In MARSSIM section 6.3 the following is stated:

> "The MDC value should be used when stating the detection capability of an instrument. Again, this value is used before any measurements are made and is used to estimate the level of activity that can be detected using a given measurement method."

And also,

> "Underestimating an MDC can have adverse consequences, especially if activity is later detected at a level above the stated MDC".

The NUREG clearly shows that the MDC is calculated under ideal conditions and at no time should be used in making determinations about detection decisions. The quotes above from MARSSIM can clearly confuse a user of the manual that the MDC should be used as a detection decision. The language in this section needs to provide a much more stringent warning about how to assess results as detectable and not use the MDC as a determining factor for detectability.

Editorial Suggestions

b) The discussions of MDC and minimum detectable count rate (MDCR) really should be reversed so that the development of measurement to result follows more logically.
c) No guidance is provided on how to determine the value ascribed to "surveyor efficiency" with regards to determining an MDCR or an MDC. In the examples provided the value of 0.5 is given to this factor. It appears from the context of the document that this value is arbitrary based on what the project team deems acceptable. Lower values of the surveyor efficiency yield higher values of MDC and MDCR and thus are non-conservative. There should be some guidance on how exactly to evaluate this term as it has a significant effect on the detection process.
d) Table 6.8 describes the hand-held instrument for analysis of smear samples as,

> "Only measurement technique for assessing removable radioactive material"

This should likely be reworded as follows,

> "Easily transportable measurement technique for assessing removable radioactive material"

Table 6.8 under Laboratory Analysis of smear samples should also be reworded as follows:

> "Only measurement technique for assessing radionuclide specific gamma-emitting removable radioactive material."

e) Figure 6.2 needs some clarification as it does not show the area of protective screen.
f) Example 9 needs to state that the detection device assumed to be used for this example is a gas proportional counter, and that the efficiency noted for the

$^{90}Sr/^{90}Y$ radionuclides is a weighted average from the two different beta particle energies (0.546 and 2.28 MeV).

RESPONSE

Appendix H

This appendix provides useful basic information regarding instruments used for detection and is technically correct as far as it describes each instrument..

Editorial Suggestion

One subject that could be stated with greater clarity is the variability of response of each of the field monitoring/survey instruments when measuring different classes and energies of radiation. This is a perfect example of where uncertainty of the measurement is extremely important. There is no indication, for any of these instruments, of their measurement uncertainty at any level[1]. This is an omission that needs to be corrected.

[1] [Note: In one instance on page H-34, it is cited that "… reportedly detects uranium and thorium concentrations at 1 ppm, or 10 Bq/kg (0.3 pCi/g) for ^{238}U and 0.4 Bq/kg (0.1 pCi/g) for ^{232}Th." It is not clear whether this information comes from the manufacturer of the instrument or was calculated by the authors. The activity values for the given mass of U or Th are incorrect. For 1 ppm of ^{238}U the activity is 12.3 Bq/kg and for ^{232}Th is 4.05 Bq/kg]

Response to Charge Question 2.2

2.2. Please comment on whether the additional optional methodology for the use of Ranked Set Sampling (Appendix E) for hard-to-detect radionuclides is useful, appropriate and clear.

The Ranked Set Sampling methodology requires a close, reasonable and provable correlation between an easy-to-measure attribute of the sample (e.g., soil sample size distribution) and the activity level of a hard-to-detect radionuclide. While challenging to implement in practice, the revisions include this optional method to assist sites with designing surveys for hard-to-detect radionuclides, which can be difficult and resource intensive to implement.

RESPONSE

The concepts put forward in this Appendix for developing a means to find areas of elevated activity concentration for hard-to-detect radionuclides *do not* provide the user with reasonable direction or assurance that HTD high activity concentration areas will be properly selected. Significant information and guidance on how to perform this type of assessment is missing.

- Introduction to the Appendix

  The recommended method for initiating the RSS seems to rely on using a field sampling instrument. Appendix E states,

  > "HTD radionuclides are typically those that emit alpha or beta particles, but no gamma rays, making them hard to detect and quantify with scan measurements, especially in soil."

  When performing such a survey it is likely that the concentrations of radionuclides of concern, in most areas of the survey unit, will be lower than the $DGGL_{EMC}$. Can measurements instruments do not have low detection capabilities for alpha/beta only emitters effectively distinguish between areas of low activity? The lower the activity concentration the more likely it will be that detection *will not* occur, and that the count rates measured will be random numbers.

  Next the italicized phrase is used.

  > "Phase 1 uses professional judgment combined with a *relatively inexpensive field screening method* to rank a parameter of interest (e.g., field survey detector count rates roughly corresponding to radionuclide concentrations in soil) within *N* field screening measurement locations."

Generally speaking, the lower the cost of the field screening technique, the less likely that detection at levels at and below the $DCGL_{EMC}$ will occur.

At the very least if field measurements are to be made, they should be done in situ, collecting data over a significant enough time period to assess if there is a statistical difference in the *total counts* measured, with aliquants of field samples analyzed in a location other than in the field. While Appendix E on page hints at this-

> "The field screening method will typically involve some type of field measurement of the samples in a consistent counting geometry. For example, the field measurement might measure the response (count rate) of an appropriate survey instrument to a fixed amount of sample material"

it really needs to state this in more positive terms.

- Furthermore, the appendix states:

> "As the interest is only in the relative comparison of the field measurements, the field screening technique does not necessarily need to be calibrated to estimate the actual concentration of radioactive material. However, any field instruments used should still be operationally checked to ensure that they are operating properly and meet the applicable MQOs".

Instrument calibration must be performed and checked routinely otherwise there is no way of ensuring that even the crude comparisons being performed will be correctly interpreted. At the very least calibration check of the instrument for the radionuclide of interests' response should be verified at the beginning and end of a series of measurements.

- <u>Examples 1 to 4</u>

In Example 1, there is significant information missing in the description and the first table. These would be:

- The radionuclides that are HTD in the investigation
- The type of instrument used to make the measurements
- Were the samples removed and taken to a low background location for analysis or was the count rate determined *in situ*?
- What was the counting geometry used?
- What was either the observation time for the count rate measurement or the time to achieve accumulated counts to achieve the recorded count rate?
- What is the estimated uncertainty of each measurement?
- Were there duplicate sample counts made of any of the sample areas?

- Were there measurements made of a background reference area for comparison?

Including all of the above information would be of value to the user in assessing how they might use this technique for their own RSS process (some of these are identified later in Example 5 but should be moved forward to the first example so that the user can see up front where this example series is headed).

The second data table identifying the activity concentration in Bq/kg is based on what, laboratory radiochemical separations? Or is it just a conversion from the cpm data? If there is more than one radionuclide potentially present, do the cpm rankings really represent the potential dose consequence for a mixture of radionuclides? Here, as well, the combined standard uncertainty must be included.

The above concepts carry through into the subsequent examples in Appendix E.

In Example 2, the definition of Δ/σ in the footnotes is incorrect (it should state that Δ/σ is the gray region divided by the standard deviation of the underlying distribution).

Examples 3 and 4 are relatively easy to follow. The associated Tables with the examples should only highlight the portions of the table that apply to the examples to make it more obvious how the table values are obtained (the highlighted sections really don't assist the user on focusing in on the important information needed in the examples).

- <u>Example 5</u>

The last sentence of Example 5 states:

> "The intent is not to expect a direct correlation between counts and concentration, only the relative ranking."

This statement is incongruous to the assumption that you can rank areas of concentration based on the count rate. If the concentration is not indicated by the count rate how could you ever assume that the highest count rate represents an area of the highest probable activity concentration?

Under Item 3 in the Example 5 background information, it discusses

> "The utility of the process has been confirmed for alpha-emitters in soil and beta-emitters in soil where the βmax energy is greater than ~250 kilo-electron-volt (e.g., ^{99}Tc…"

The $E_{\beta max}$ for ^{99}Tc is only 290 keV. This means that only about 10 % of all decays will be detected. This combined with the low activity concentration expected would not make for a very effective means of assessing ranking of the elevated areas.

Editorial Suggestions

Under Item 6 in Example 5, it would be helpful to indicate how the area factors in the table were determined. There is no way to use the information in this document to determine those factors. Furthermore, there should be some reference to Appendix O in this example to show how the area factors are used to calculate the $DCGL_{EMC}$.

In Step 12 of the Example 5 analysis details, it states the following:

> "The net result is that 15 laboratory samples were required for the SRS sign test, but the requirements were increased to 45 laboratory samples to account for areas of elevated concentration of radioactive material that could reasonably be expected. This process closely parallels the more familiar required/actual scan MDC paradigm used successfully for MARSSIM soil surveys involving gamma-emitting radionuclides"

To make this example more understandable, how one comes to the conclusion that 45 samples vice 15 samples are required would be extremely helpful. As written, the user is left to wonder what are the derivations of these decisions.

Response to SAB Charge Question 2.3

2.0 Does MARSSIM, Revision 2 provide useful, appropriate and clear examples and descriptions of technical approaches to implementing surveys and the statistics by which they are interpreted?

2.3 Please comment on whether the new and additional examples provided in Chapter 5 are useful, appropriate and clear.

RESPONSE

General Utility of Examples

Overall, the examples are useful and provide a means for users to see how the manual guidance can be implemented for practical situations. However, each example could use additional detail on calculations and use of the tables cited in the text. It cannot be assumed that the user will be well versed in either sampling theory or use of the statistical models employed in the examples.

Editorial Suggestions

Example 1in Rev 2 is identical to the Rev 1 Example Scoping Survey Checklist.

The section on characterization surveys provides minor guidance on the scoping of contamination in other matrices besides soils (e.g., ground water, surface water and sediments). While such information is important to assessing how site contamination is impacted, MARSSIM of itself does not provide guidance on how to use this additional information to structure site clean up or future site surveys. Example 2 in Rev 2 is fundamentally the same as Example Characterization Survey Checklist in Rev 1.

Example 3 in Rev 2 is identical to the Rev 1 Remedial Action Support Survey Checklist

Starting with the same section on Final Status Survey, the two document descriptions and examples diverge significantly. Both contain flow charts identifying the flow paths for determination of survey plan, measurement locations and needs assessment. Rev 2 however goes into significant details about selecting the appropriate scenario and an expanded discussion of the gray region and the selection of a lower bound of the gray region.

The first ‘example’ in Rev 1 starts on page 5-29. That example is similar to Example 4 in Rev 2, both of which describe a Scenario A situation. The description of the process for selecting the number of samples in the reference and test area is clearer in Rev 2 than in Rev 1.

The next example in Rev 1 is at page 5-33 and is similar to Example 5 in Rev 2. The description of the process for selecting the number of samples in the reference and test area is clearer in Rev 2 than in Rev 1.

Example 6 in Rev 2 is new. It provides another example for a Scenario A decision on the number of samples to be selected. The example clearly identifies the input decisions needed by the project team on TypeI/II decision errors, the $DCGL_w$ and the LBGR value. It also clearly shows the calculation involved and the use of the Sign Test table. What would be very helpful to users would be to demonstrate that as the significance of the Type I or II errors increase, from let's say 0.05 to 0.01, that the number of samples required significantly increases. Then discuss that the Type I and Type II decision error rates are based on the project DQO values. This type of discussion then ties the decision made using the data obtained back to the original MQOs and DQOs.

The Examples 7 and 8 provided in Section 5.3.5 demonstrate how to use the information discussed in 5.3.5. however, there are assumptions made for the initial calculation of the number of samples to be taken that should be stated in the example so that how the numbers were determined is clear to the reader. For instance, in Example 7, where did the value 27 come from? There is a need to show this calculation and the use of Table 5.3 in the example.

Example 8 provides good contrast to the scenario depicted in Example 7 and lets the user see how MARSSIM can be an iterative process. A minor item that should be corrected in Example 8 is the following:

> "The grid area encompassed by a triangular sampling pattern of 10 m is approximately ~~86.6~~ 99.7 m^2, as calculated using Equation 5-3:"

Example 9 is used to demonstrate a random sampling pattern for a Class 3 area. It incorrectly references Table I.11 as it should be Table I.12. This type of example is beneficial to the user providing guidance on random selection of the sampling locations. However, it is unclear how Table I.12 is used to select the locations depicted on the diagram. The method used should be explained in some detail.

Example 10 provides a good means of showing the triangular grid sampling pattern based on the equations used in the chapter. The random start coordinates are stated as a foregone fact without any development. The stated location relates to Table I.12 but it is not obvious how you identify that random start point.

Example 11 identifies a sample checklist for the FSS. It provides additional details not included in the Rev 1 version. Checklists such as these provide users with both goals to attain during the remediation and final survey status as well as serving as a final quality check that all items have been considered.

Mr. Dennis Quinn

3. *Is the information in MARSSIM, Revision 2 clear, understandable, and presented in a logical sequence? How can the presentation and content of material be modified to improve the understandability of the manual?*
   3.1. *Please comment on the revised description of how to set the Lower Bound of the Grey Region (LBGR) and its likely effectiveness in encouraging users to rely on site-specific information for doing so (Chapter 4 and Section 5.3).*

   One of the critical decisions made during site survey design under MARSSIM Scenario A is to set a value for the LBGR. Twenty years of training and review of survey plans have shown that this concept is not well understood by users, and that users tend to implement the standard rule of thumb of setting the LBGR to 50% of the DCGLw. This rule of thumb was provided in MARSSIM, Revision 1, for use only when additional information was not available. A poorly chosen value for the LBGR can affect the power of a survey resulting in unnecessary use of resources or a higher chance of failing a survey unit that meets the release criteria.

   In Scenario A, the LBGR should be set equal to a conservative estimate of the average concentration remaining in the survey unit. This information is typically available from historical site information, or a scoping or characterization survey if the survey unit is un-remediated, or the remedial action survey if the site has been remediated. The purpose of the revisions is to describe this concept in plain language, moving away from a statistics terminology description of the concept.

The sections on determining LBGR and determination of the relative shift (4.12.3, 5.3.3 and 5.3.4) were understandable, and the examples were clear. The use of plain language vs. statistics terminology made for easier comprehension. I expect that with these improvements, the concepts of LBGR and determination of relative shift would be better understood, then the user would be more likely to use site data instead of a rule of thumb.

3.2. Please comment on whether avoiding the use of the term "area factor" improves understandability of the elevated measurement comparison concept (Section 8.6.1).

Area factors, which are simply the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria, should be based on site-specific modeling or calculations. Due to the misapplication of published area factors from the literature and to provide focus on the need for development of site-specific EMC criteria, MARSSIM, Revision 2 avoids the use of the term area factor. In addition, lessons learned from training MARSSIM show that describing the EMC

> *concept in descriptive language, rather than by defining additional terminology, seems to improve understandability of the concept.*

The term area factor was adequately described in Marssim Rev. 1. It is described in detail in Appendix O of Rev. 2, but I believe it is not adequately described elsewhere. The need to use site-specific modeling was included in Rev. 1, but apparently that was not well understood by some users. My suggestion would be to include the discussion of Area Factor in Chapter 8 and include clear examples and additional precautionary statements. Perhaps do not include the Tables from Rev. 1 (5.6 and 5.7), which could lead to inappropriate use of those tables.

3.3 Please comment on the effectiveness of the new organization of Chapter 4 (Considerations for Planning Surveys) to improve the understandability of the Chapter.

> *Earlier reviews of Chapter 4 provided evidence that the fundamental organization of Chapter 4 made it difficult to find and understand vital information. After discussing the challenge with experts in training and explaining the material, Chapter 4 was completely rewritten or reorganized in an attempt to improve understandability without changing the fundamental purpose of or material in the Chapter. In an effort to streamline the presentation of material in Chapter 4, some information was moved to Appendix O.*

The revised Chapter 4 in Rev. 2 appears to be well organized and more comprehensive than Chapter 4 in Rev. 1. It would be helpful to add information on EMC if not included elsewhere.

3.4. Please comment on the effectiveness of moving derivations from Chapter 5 to Appendix O to improve the understandability of the Chapter.

> *In an effort to streamline the presentation of material in Chapter 5, some derivations of key concepts were moved to Appendix O.*

The changes to Chapter 5 did help somewhat to improve the understandability of the chapter. The following are some of the improved aspects of the new Chapter 5:

- The addition of Figure 5.4, the flowchart for the “Process for Designing an Integrated Survey Plan for Final Status Survey.”
- Improved cross-referencing to other chapters and appendices.
- Improved format of Example Survey Checklists.

Dr. Daniel Stram

1.1 Are the revisions to MARSSIM concepts and methodologies technically appropriate, useful and clear, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

Overall, the statistical methods are appropriate and useful, but at times the description is unclear. Specific comments are below.

Section 5.3.6.1 gives a simple formula (Eq. 5-10) for the fraction of a survey unit that should be scanned for Class 2 and Class 3 areas using instrumentation suitable for scan-only surveys. For Class 1 regions 100% of the survey unit should be scanned. This rule is clear although a rationale behind Eq. 5-10 is lacking. For Class 1 areas the main issue regarding measurement uncertainty is bias since scan-only methods must quantify as well as detect radiation. While scanning the entire site with perhaps thousands of scan data points minimizes certain types of sampling error, calibration must be performed to quantify and remove systematic errors. Rescanning certain areas, while helpful, is in general insufficient to quantify measurement uncertainty, comparison to a “gold standard” laboratory measurement is required. These concepts are rather nebulous, for example the third point (lines 23-24, page 5-43) indicates only that a correlation must be established between scan-only and laboratory results. This is too vague; if the scan-only systems are to be used for quantification then a very strong relationship between the two should be established.

Section 8.5: With a scan-only survey, in order to demonstrate compliance with the release criteria an upper (1-alpha) x 100% confidence limit (UCL) for the mean is constructed (Scenario A), where alpha is the type 1 error controlled for by the study design; the release criteria are regarded as met if this UCL is below the upper bound of the gray region. (for Scenario B beta replaces alpha). The confidence interval recommended is based on the Chebyshev inequality and can be very conservative, meaning that it tends to accept the alternative hypothesis when it is false much less often than specified by alpha, which results in loss of power. The reason that this “UCL test” is proposed for scan-only surveys but not for others seems to be simply that there are so many individual results taken is that the loss of power is made up for by very large n (in Eq. 8.3).

Even the UCL test does not guard against other sorts of uncertainty, if there is non-negligible measurement uncertainty (as seen compared to laboratory measurements) then this can result in loss of control of type I error probability, i.e. by overestimating the relative shift. It is recommended that the scan MDC be less than ½ of the $DCGL_W$, and that other MPOs for scan-only measurement uncertainty be met. This criterion seems reasonable in the context of this report, although it shouldn’t be taken as a default when no calibration data exists. The upper bound of the gray region certainly needs to reflect measurement uncertainty, despite the additional complexity in ascertaining appropriate values.

It is stated in section 4.8.6 that in general FSS is not to be based on scan-only methods. This recommendation is not evident in Fig. 5.4 or in the discussion in Section 5.3.6 and would imply

that the measurement uncertainty is greater for scan-only than for other methods. If this is not a fair statement, then this should be weakened.

1.2 **Please comment on the inclusion and proposed implementation of Scenario B (Chapter 4, Section 5.3, and Chapter 8).**

Overall, once one has gotten used to the terminology the proposal for including Scenario B as a recommended approach when the DCGL is close to zero (no added contamination) is clear and appropriate. While it is possible to reformulate Scenario A to deal with low DCGL (by replacing the UBGR with a detection limit) it is more natural to revert to Scenario A, as long as power is well controlled in the analysis.

Is it appropriate to recommend that Scenario B be used only those situations when Scenario A is not feasible?

I find that recommendation of Scenario A as the default approach for designing a study is reasonable and agree that it is reasonable to only recommend scenario B when the criteria being tested against is equivalent to there being no residual radioactivity left after remediation.

Most people would agree that an inadequately designed study should not be used to accept that mean concentration is below a certain level (the DCGL), justifying choice of Scenario A. However, if that concentration level is too close to zero then the relative shift (Δ/σ) will be too small to give a reasonable sample size. just switching scenarios from A to B is not sufficient, one needs to do something to increase the relative shift; in sample size calculations a new UBGR is required, one that can be reliably distinguished from the very small DCGL. In Scenario B the DCGL remains small, functioning as the LBGR (and renamed as the AL) while the UBGR is now set to the discrimination limit (DL) which is defined as the concentration that can be distinguished from the AL. On the other hand, if we were to want to keep Scenario A terminology then UBGR must be enlarged by increasing the DCGL. So long as this increase in the UBGR under scenario A equals the DL under Scenario B then, for the purpose of power calculations, there is no difference since the relative shift remains the same, except that type 1 and type 2 errors switch.

While in this sense equivalent, the practical issue that comes up when applying Scenario B, is that while p-values are evaluated using the estimate, $\hat{\sigma}$, of σ, computed from the observed data – type 2 errors, β,only come up in the design phase. If we over-estimated the relative shift in designing the study, this will not matter in computing a valid p-value, but it will mean that type 2 errors have been underestimated in the design phase (i.e. the power has been over-estimated). In Scenario B it is crucial that type 2 errors have been properly controlled in order to confirm that the strength of the evidence in favor of the null hypothesis that the mean concentration is below the DL; this is why retrospective power analysis is recommended when using scenario B. The key to retrospective power analysis is comparing the observed relative shift ($\Delta/\hat{\sigma}$) using the value, $\hat{\sigma}$, of σ estimated during the study, to the Δ/σ used in the design phase, as further described in appendix M.

Are methods for considering background variability in assessing whether the site is indistinguishable from background reasonable and technically accurate? Is the inclusion and proposed implementation of added requirements for retrospective power analysis and the Quantile Test while using Scenario B technically appropriate and discussed adequately and clearly?

The MARSSIM document adopts statistical methods that test for both changes in location (with emphasis on shifts in the median) and changes in the upper portion of the distribution, both are important for the evaluation of residual radioactivity. The main issues in the analysis of data under Scenario B are: (1) whether appropriate statistical tests have been used for data analysis and (2) whether sample size and power had been appropriately calculated in the design phase, i.e., had the relative shift been properly estimated. The recommended, nonparametric, tests under Scenario B, are the Wilcoxon rank sum test and Quantile test. Down on the list are parametric tests (one and two sample t-tests, which require). An argument may be raised against nonparametric tests in general, since they mostly are used to reduce the influence of outliers (e.g., skewness and other failures to be normally distributed) on the estimation of location parameters. Both the Sign and the Wilcoxon Rank Sum (WRS) test are testing for a shift in the median rather than the mean and as such are very robust to the influence of outliers. However, possibly because risk models are generally linear in radiation exposure, the focus in MARSSIM, is stated to be on controlling mean exposure. Because of this a combination of the Wilcoxon test and the Quantile test is recommended in scenario B, as well as inspection of the sample mean compared to the median. The Quantile test is sensitive to shifts that only affect some of the data (i.e., the upper portion). Furthermore, values that exceed an investigational limit are to be examined more closely, despite the outcome of the non-parametric tests (see Section 8.6.1).

The description of the use of the WRS test is clear and can be readily implemented in spreadsheet format (and almost all general-use statistical packages provide this test). Section 8.4.2 indicates that to apply the WRS test for scenario B that first the AL (*here equal to zero?*) is subtracted from the survey unit measurements and then what boils down to a one-sided Wilcoxon rank sum test is performed on the adjusted measurements. Otherwise, the Quantile test is performed (again on the adjusted survey unit measurements, and if this test also is not rejected then the survey unit is judged as not exceeding the release criteria. Table 5-2 is provided for power calculations for the WRS test based on equation O-1 in the appendix and can be applied both prospectively and retrospectively. No power calculations are given for the quantile test. This seems like a problem for Scenario B since retrospective power calculations are considered essential before accepting null results. The Quantile test seems to be used more as a safety valve, than a fully characterized statistical procedure, guarding only against the worst possible scenarios involving non-normal tail distributions in the reference or survey areas (or both). Other possibilities come to mind such as using the Quantile test first, and if null, following this with t-tests, however given the difficulties of the problem of estimating differences in means for skewed data, the combination of a Wilcoxon and Quantile test appears to be a reasonable, if not a perfect choice.

Under hypothesis testing in MARSSIM, Scenario B is defined as assuming that the survey unit meets the release criteria unless proven otherwise, and its use was discouraged in MARSSIM, Revision 1. However, this is the only viable option for sites where the criterion is effectively "no added radioactivity" or "indistinguishable from background".

See above remarks, agree that Scenario B should be preferred only when the release criterion is close to or at zero.

In Scenario B, the Lower Bound of the Grey Region (LBGR) is often set to zero, but the document allows use of a non-zero LBGR that considers background variability in determining whether the survey unit is indistinguishable from background

Example 9 on page 8-35 describes the issue of background variability on mean concentration levels from reference site to reference site. It does seem to be appropriate to account for this if it really is an important source of variability, however this may require many additional measurements to carry out consistently. Why are 3 standard deviations of the variance of the random effect chosen? This is justified in NUREG-1505 by looking at the probability that an arbitrarily distributed random variable is more than *t* standard deviations away from its mean using Chebyshev's inequality

$$\Pr(|(X-\mu)| \geq t\sigma) \leq 1/t\char`^2.$$

With t equal to 3 the probability is less than 1/9 = 0.11. (Actually a one-sided bound is more relevant to this problem and is given as $(\text{Prob}(X\text{-}\mu \geq t\sigma) < 1/(1+t^2))$ or 0.10). This bound can be very weak, for a normally distributed random variable the one-sided probability is just 0.002. Recommend using something like $t=3/2\sigma$ (1-tail error normal probability of 0.89). Here σ is the standard deviation of the between reference site random effect estimated in a one-way random effects analysis of variance.

Since Scenario B assumes that the site meets the release criteria, there is a risk that the survey unit will pass simply because the survey did not have sufficient rigor. To guard against that, the revisions require that when using Scenario B, the survey unit must perform a retrospective power analysis to prove the survey has sufficient statistical power to detect a survey unit that should not have passed

I agree that retrospective power analysis is a key component of the evaluation of the evidence in favor of the null hypothesis. Power for the WRS is a knotty problem but there are approximations (Eq. O-1) that are helpful in evaluating power retrospectively as well as prospectively as needed. No method of performing a retrospective power analysis is suggested for the Quantile test.

The non-parametric tests included in MARSSIM test the median instead of the mean. The release criteria are typically expressed as the mean. To guard against Scenario B situations where the median will pass but the mean won't (this can occur in sample data distributions with a long tail in the higher concentration range), Revision 2 also requires that when using Scenario B, the survey unit must pass a quantile test to guard against excessive skewness

Both of these recommendations seem appropriate see above. One possible concern is that if the quantile test is very low power then giving it half of the type 1 error will weaken the power of the overall WRS + quantile test. This could be considered further perhaps by a simulation study (although the impact is likely to be small).

Other comments

Under Scenario B the gray region is defined in the glossary as having lower bound equal to the $DCGL_w$ and upper bound equal to the discrimination level (DL). However, throughout the main document (e.g., Figures 5.7 & 5.8, pages 8-28 - 8.30, etc.) the lower bound is termed the AL (action level), is the AL generally taken to equal the $DCGL_w$ in Scenario B?

1.3 Is the proposed implementation of the of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)?

It would be helpful to make a further clarification / distinction between "shared" and "unshared" measurement uncertainty most relevant to scan-only surveys. Unshared uncertainties (such as the inherent variability of Poisson counts around the mean) will tend to average out to zero as the number of measurements is increased. Ignoring unshared measurement uncertainties in sample size calculations will lead to studies that are too small since σ is under-estimated and hence the relative shift, Δ/σ, over-estimated. In examining the data, it should be evident however that the σ used in the design stage was too small so that confidence intervals around the mean (using estimated σ), will be appropriate, and retrospective power calculations will show that power was miscalculated during study design.

The situation is worse for shared uncertainty (i.e. bias). For example, in a scan-only survey, if different operators scan at different speeds and different distances from the surface then this can lead to operator differences in recorded values which are shared between all the individual scan points recorded by that operator and which do not tend to average out to zero even if the number of scan points is increased greatly. If these operator differences are ignored, then this can greatly affect performance of the statistical tests which all assume that results for individual scan points are independent estimates of the true mean (or median) for a survey. Completely shared operator uncertainties (if the entire survey site is scanned by one operator) may (if not controlled for by having the same operator scan all reference and survey sites) , give too small an estimate of σ and hence too large an estimate of the relative shift, but these uncertainties may be invisible in the data generated by the survey, therefore retrospective power analysis may not indicate that the study design was flawed, and analysis of the data can give confidence intervals (even those based on Chebyshev's inequality) that are too narrow. Partly shared measurement uncertainties have intermediate effects, as when there are multiple operators, but operator is not controlled for in the analysis. Note that the terminology, "shared" and "unshared" measurement uncertainties includes the previous (Revision 1) distinction between "systematic" and "random" errors as well as "bias" and "precision" in Revision 2, as special cases.

From these considerations I recommend that, despite the difficulties in fully characterizing measurement uncertainty, that its characterization remain a part of the MQO. Consideration of whether measurement uncertainties are independent or shared should be a part of the MQO and if shared uncertainties (bias) are non-trivial, plans to mitigate their effects (e.g., by calibration to laboratory readings and controlling for operator differences in the design of the study) should be included.

3.1 Is the information in MARSSIM, Revision 2 clear, understandable and presented in a logical sequence? How can the presentation and content of material be modified to improve the understandability of the manual?

Overall, the material, once the reader has gotten used to the terminology, the presentation is understandable and presented in a logical sequence. My main concern is with the definition of LBGR, AL and DL for Scenario B (see below)

Please comment on the revised description of how to set the Lower Bound of the Grey Region (LBGR) and its likely effectiveness in encouraging users to rely on site-specific information for doing so (Chapter 4 and Section 5.3).

Throughout the MARSSIM manual the relative shift, Δ/σ ,which determines sample size (for fixed α and β error probabilities) depends on three quantities, the LBGR, the UBGR, and the standard deviation of the measurements σ, i.e.

$$\Delta/\sigma = (\text{UBGR} - \text{LBGR})/\sigma.$$

For Scenario A the lower bound of the gray Region (LBGR) is defined quite clearly and in plain language as equal to or slightly greater than an estimate of residual radioactive material remaining in the survey unit. Since for Scenario A the upper bound of the gray region (UBGR) is also clearly defined (as equal to the DCGLw), and since the standard deviation, σ, is also a well-defined quantity, then so is the relative shift. However, the source of information about the UBGR and the standard deviation is somewhat less clear, the listed sources (page 5-29, lines 20-

23), i.e. non-judgment survey data (e.g., scoping or characterization survey data for un-remediated survey units or RAS surveys for remediated survey units), while containing some relevant information some of the time, are likely to provide only a few, data points for a given unit and these will often be non-representative of the status of the reference and/or survey areas. It is probably just human nature that seeks to replace a decision that requires some thinking (about how best to use scarce preliminary data) with a hard and fast rule (set the LBGR to 50% of the DCGLw), even if it results in a more costly survey, or inadequate power. This rule of thumb does not address σ but it seems to be possible (based on my reading of the bottom of page 5-29 to the top of 5.30) that a coefficient of variation of 30% of the DCGLw, might be acceptable "based on experience", and would give a relative shift of 1.67 which also may be acceptable (page 5-30 line 20) but probably leads to larger sample sizes than really required in most cases. My question for the MARSSIM authors is in what fraction of the actual cases is truly informative data ignored in this sample size determination? Given the other places where conservative assumptions are employed in these calculations is a rule of thumb similar to the above really that bad?

It is Scenario B where both the LBGR and UBGR need further clarification. In the glossary the LBGR for Scenario B is taken to be the DCGLw, in the body of the manual however it is set to a not well-defined action level (AL) that sounds a lot like a DCGL, and the UBGR is called very vaguely the discrimination limit defined in statistical terms as the value that can be "readily distinguished from the AL" (page 2-12, line 18). This is somewhat circular since such a limit depends upon sample size. I recognize that such ambiguity is inherent to these calculations, in my experience it is often helpful to present several power curves as a function of the DL for a series of sample sizes and judging whether the DL for a "reasonable" study size is not unrealistically large in relation to typical requirements. It would be helpful if an example case of determining the DL be included as an example in the body of the manual.

Mr. Zoltan Szabo

1) Are the revisions to MARSSIM concepts and methodologies *technically appropriate, useful and clear*, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces

1.3) Is the proposed implementation of the of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)? *Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives? Is the method appropriate and practical for both laboratory and field (including scan) measurements?* Please comment on the concerns of stakeholders that calculating measurement uncertainty for field measurements makes the survey process difficult to implement. In addition, please comment on whether recommendations provided by NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).

The concept of MQOs as a subset of Data Quality Objectives (DQOs) originated after publication of MARSSIM, Revision 1. The use of MQOs ensures that each measurement taken is of sufficient quality to be used as part of the survey design. These MQOs include many familiar Data Quality Indicators, which were included in MARSSIM, Revision 1, such as range, specificity, ruggedness and detection capability, typically represented as MDC/MDA. However, the older Data Quality Indicators of bias and precision have been captured by a new MQO: measurement uncertainty, with bias indicating systematic uncertainty and precision indicating random uncertainty. The International Organization for Standardization published the Guide to Uncertainty in Measurement in 1995. The National Institute for Standards and Technology (NIST) published Technical Note 1297: Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results (GUM) in 1994, which provided guidance to the federal government to incorporate measurement uncertainty into their procedures. As a result, subsequent MAR-series documents MARLAP and MARSAME included information on the use of measurement uncertainty.

MARSSIM, Revision 1, indicated that the greater source of error for a survey was typically found in the sampling design, not in the measurements themselves, and as a result, did not emphasize concerns regarding measurement uncertainty. However, with the inclusion of scan-only surveys, the sampling design error decreases significantly as a greater percentage of the survey unit is covered. Consequently, the measurement error becomes critical, and thus the more quantitative method of assessing and controlling measurement uncertainty similarly becomes critical. Stakeholders have expressed concerns that calculating measurement uncertainty, specifically for field measurements, makes the survey process

difficult to implement. The MARSSIM Workgroup agreed to include the MQO for measurement uncertainty and investigate future tools to make process easier.

The American National Standards Institute/Institute of Electrical and Electronics Engineers (ANSI/IEEE) standard N42.23 recommends that interpretation of survey data involving environmental media, "such as soil, sediments, concrete and water, should not use the MDC/MDA to evaluate measurement results, and instead recommends use of the decision level or considering the confidence interval for the measurement result." The authors of MARSSIM, Revision 1, understood that for cases when the decision to be made concerns the mean of a population that is represented by multiple measurements, detection criteria based on the MDC/MDA may not be sufficient and a somewhat more stringent requirement was needed. To meet this need, they introduced an additional requirement that the MDC/MDA should be set at 10-50% of the action level. This predated the concept of measurement quantifiability (as considered in MARLAP and ANSI/IEEE N42.23), but it results in comparable constraints on a Minimum Quantification Concentration (See MARSAME Section 7.6 for further discussion.) To minimize changes to current practice, the original MARSSIM requirement is left as is in Revision 2

Response: Overall, I found a split decision regarding the adequacy and accuracy of descriptions for Chapter 4 and Appendix D, regarding the concept of measurement uncertainty and its inclusion in study designs going forward with implementation of this version of MARSSIM Guidance. Reading Chapter 4, I was did not find the adequacy of descriptions had been met, especially in the first half of the Chapter that needed to establish the necessity for, and the details of, the use of MQOs as part of the Data Quality Objectives (DQOs) planning and implementation process. Many sections consisted primarily of bullets and very short summaries. The underlying logic for the planning process and succession of steps was often missing or unclear. The Chapter left me with the sense that the question 1.3. had not adequately been addressed. Regarding the Question: ***Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives?*** –Chapter 4 certainly indicates that YES, it is, but the Chapter is short of the details. Regarding the Question: ***Is the method appropriate and practical for both laboratory and field (including scan) measurements***? –I simply found minimal discussion of either appropriateness or practicality for especially field measurements, none that were detailed enough to allow in-depth assessment. The first half of Chapter 4 read almost like the Executive Summary of the process, but even then, vague and not directly to the point. I hoped the Example Problems at the end would be clarifying, but they were not. They were too broad in scope and required reading several additional Chapters and Appendixes; they needed to be focused and concise that directly addressed very narrowly (1) the topics of the Chapter Title "CONSIDERATIONS FOR PLANNING SURVEYS" and (2) the second of the above questions: ***Is the method (proposed calculation of measurement uncertainty) appropriate and practical for both laboratory and field measurements***?. "Practical" was never addressed. The Example Problems would be more helpful if they narrowly illustrated how the method (proposed calculation of measurement uncertainty) could be effectively implemented for field measurements; and also, if they gave examples where the project personnel considered in preliminary manner how to evaluate the difference resulting from using the proposed calculation of measurement uncertainty versus the previous (MARSSIM v1, MDC/MDA set at 10-50% of the action level) technique. Another Example showing an evaluation in interpreting the results and reaching a final decision accounting for possible differences in outcome depending upon approaches to Data Quality Management would be useful. Missing throughout Chapter 4 are appropriate call-ins to details that are presented in Appendix D that would strongly help in using Chapter 4, which as of now, is more of a lengthy Summary of concepts. Rather than simply calling in the entire Appendix D near the beginning, specific detailed enumerated sections of Appendix D (say, Appendix D.3.4 or AppendixD.4.2.4) might be called in to point the reader to the additional details regarding topics in Chapter 4. The literature cited is appropriate, but is mostly contained in Appendix D, and consists of further manuals that are also often hundreds of pages long. Perhaps citing specific chapters from within these Manuals that are focused on specific topics of concern (such as Measurement Quality Objectives) as opposed to just the entire Manual, may help the readers who wish to obtain additional information for specific considerations regarding the narrower uncertainty and MQO topics.

Regarding Chapter 4, while it is stated that Stakeholders have expressed concerns that calculating measurement uncertainty, specifically for field measurements, makes the survey process difficult to implement, but specific examples are not given or discussed/evaluated. The discussion of either appropriateness or practicality of measurement uncertainty approach for especially field measurements is minimal. It is suggested that the main criteria that could/would be used for choosing the principal MQO methods implemented for the study need to be listed perhaps in a Table (if various uncertainty methods remain allowable, as it appears). This Table & discussion should appear in Chapter 4, not the Appendix D. If the uncertainty process is too difficult to implement in some limited circumstances, a summary table of various approaches or considerations that might be employed instead to limit effects for data analysis and decision making may be a useful addition. In addition, perhaps revise the Example Problems geared to show in more detail how different implementation approaches might reach difference in interpretations and decisions.

Finally, while the MARSSIM Workgroup mentions the need to investigate future tools to make the process easier, there is simply no mention of this topic, nor any guidance as to what those tools might consider. Even if only a paragraph, a mention is suggested. Many useful guidance documents are cited in the Appendix D that are not mentioned in Chapter 4, but Chapter 4 would benefit from a call-in pointing to them.

In contrast, the presentation in Appendix D was detailed, logical, and presented information regarding the complexity of the issues at hand, including a far more detailed discussion of the Planning Process, and more commentary regarding the MQO, its importance, and more details. Examples were provided that were narrow enough in scope to help clarify individual points. Sections in Appendix D were also somewhat more self-contained than Chapter 4 overall, with in the first half especially focusing primarily on the planning procedure, MQO, uncertainty, and the sequence of steps to complete the survey successfully. The Appendix D is subdivided into 4 sections as follows: 1) The Planning Phase, 2) The Implementation Phase, 3) The Assessment Phase, and 4) Data Verification and Validation. For each Section and subsection, the key points are clearly discussed, planning steps are clearly listed and detailed, and the logic behind them are clearly described. Each subsection is mostly self-contained and focused on the details. The Appendix D also contains numerous simple, informative, organized figures and tables that summarize in easily understandable way the key points and steps.

Regarding the Question: ***Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives?*** –Appendix D certainly indicates YES, and the details are adequately discussed. Regarding the Question: ***Is the method appropriate and practical for both laboratory and field (including scan) measurements***? – discussion of the appropriateness is detailed and adequate; somewhat still missing is a detailed discussion of the practicality for especially field measurements. If the current structure is maintained (Chapter 4, more summary-type, and Appendix D with all the details), then a Table with various approaches or considerations of implementation of the method for troublesome field measurements might be beneficial, certainly in Appendix D, but maybe in Chapter 4. Appendix D.1 The Planning Phase is much more comprehensive and clear than the same Section in Chapter 4 – perhaps the Appendix D.1 Planning Phase should be placed to begin Chapter 4 – then nuts and bolts of the measurement difficulties with certain field techniques and issues as to how this affects MQO could be expanded in Appendix D.

Ultimately, the procedure choices that are most reasonable are the ones that, given the possible combination of less than ideal data and measurement difficulties with certain field techniques, will result in demonstrably meeting MQO so that the data can be successfully used to assess project objectives. MQO needs be met so that use of the data can lead to high degree of confidence for the decision endpoint. Uncertainty metrics that clearly define MQO with calculations that are transparent best achieve this objective. The proposed calculation of measurement uncertainty ought be used even if the difference in using the previous technique is relatively modest and does not lead to a change in the decision-level metric. The most thorough and transparent MQO and uncertainty analysis is reasonable and should be simplified only in unusual cases. One would expect that the Most Reliabe study design and MQO be incorporated into any of these studies to avoid having to repeat the study. I would suggest adding the details discussed above to indicate examples of the uncertainty limitations in achieving robust decisions, and examples of where the uncertainty process may be difficult to achieve. These could be part of Appendix D.

While addition of details is suggested for Chapter 4, and further guidance and examples to Appendix D, some minor editorial comments relating to especially Chapter 4 are also provided.

4-2L21-30-38.4.3L1-2. Data Quality Objectives Process. & D-4.Figure1.

Wording differs. Headers should be the same. Explanatory terms can be expanded upon to provide additional clarity.

4-4L21. Samples are authentic?

Are reliable, dependable, trustworthy, valid?

4-5L6-8. Information on selecting the number and type of QC measurements for a specific project are provided in **Section 3.4;** Tables 4, 5, and 6 of the UFP-QAPP Part 1; and Worksheet 28 of the UFP-QAPP Part 2A.

Why not have some of these included in Appendix D, or even in Chapter 4? This part of the discussion is central to the entire chapter and the whole discussion regarding MQO; why make it difficult for the reader to view the most essential of these tools (such as Tables)?

4-5L16.measurement performance criteria

Defined where? Should be defined there, even if defined elsewhere.

4-28L22-38.4.8.3.1 Scanning Measurements & 4.8.3.2 Direct Measurements

Calibration of instruments, tracking and requirement thereof, are not mentioned in these sections (although calibration has its own section).

4-28L34.”scanning speed and operator training”

These disparate topics should be broken into separate bullets with minor explanations.

4-31L37.realistic or conservative estimate of the MDC be used instead of an optimistic estimate

This sentence should be more specific. Suggest Use of 95% Type I error?

4-53.L6.UCL=4.6+1.96 × 1.7=8.0

While basic mathematic operon rules apply, use of parentheses is preferable

4-55.L18. alpha of 0.05 and beta of 0.10

Why not use Type-I error alpha and Type-II error beta?

D.D-23.L8-11. More information on Scenario B can be found in the NRC draft report NUREG-1505

Why not discuss implications of Type I error alpha for Scenario B? While the citations are thorough, the concept is critical at this point of the discussion, and expansion (brief summary) is reasonable.

1.4. Is the discussion of survey requirements for areas of elevated activity technically accurate, appropriate and clear? In particular, please comment on the decision to maintain the use of the unity rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4). ***Are there suggested alternatives to the use of the unity rule?***

While modeling is outside the scope of MARSSIM, depending on the modeling tool or methodology used to develop release criteria, the use of the Unity Rule for multiple areas of elevated activity in a single survey unit can lead to unrealistic or overly conservative assumptions. For example, the models may assume that the receptor is located directly above each area of elevated activity and stays there for the duration of their exposure period. This physically cannot occur in cases where there is more than one area of elevated activity per survey unit and results in concerns that this will cause an over-estimate of dose or risk, leading to an emphasis on remediating areas of elevated activity that don't incur additional significant dose or risk to receptors.

MARSSIM, Revision 2, does not change recommendations for the use of the unity rule, but emphasizes assessing whether criteria for areas of elevated activity apply to survey units, and when they do, using a common sense approach to applying these criteria, keeping in mind the limitations of the unity rule described above for multiple areas of elevated activity

Response: Yes, the discussion of survey requirements for areas of elevated activity is technically accurate, though perhaps could use more detail in the text to increase clarity. Figure 5.6 describes the needed steps in sequence making the text easy to follow. While the use of the Unity Rule for multiple areas of elevated activity can lead to overly conservative assumptions, but as the initial default position, it is preferable to one that is unreasonably relaxed and increases the likelihood of inappropriate site release. It is also agreed that it is always acceptable and conservative to assume the smallest area factor possible. Chapter 5 & 8 could use additional call-ins to the appropriate sections in the Appendixes to facilitate the clarification of details in the discussion.

O7.L25. “When applicable, As Low As Reasonably Achievable (ALARA) criteria should be considered”. The statement should be expanded to provide an example.

Examples 7 & 8: also provide the calculation for the number of chosen samples. While this is covered in Examples 4 & 5, it is a benefit to walk through all the steps to simulate a real case.

In Example 8, the purpose for the following statement is unclear: “The grid area encompassed by a triangular sampling pattern of 10 m is approximately 86.6 m2, as calculated using **Equation 5-3: “** The very next line completes the calculation, showing it to be 99.1 m2. Unclear, why is the estimate needed?

As a final comment, the reviewer conducted a limited literature search regarding the question -- *Are there suggested alternatives to the use of the unity rule* –and had no success in finding a well-documented alternative.

1.5. Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles *technically sound and appropriate, and is the description accurate?* In particular, please comment on the rule-of-thumb for determining when use of MARSSIM may not be appropriate for survey units containing discrete radioactive particles (Section 4.12.8 and Appendix O.5).

Discrete radioactive particles have an extremely small size and contain enough activity that survey units containing discrete radioactive particles generate impractical survey designs under MARSSIM. Over MARSSIM’s twenty-year history, several sites have attempted to utilize MARSSIM to address discrete radioactive particles, with predictably extreme survey designs as a result. In addition to being impractical, designs for discrete radioactive particles violate some of the assumptions commonly made during modeling, which includes parameters based on an areal source of radioactive material, e.g., length of the area of the elevated activity in the direction of overland flow. While modeling is outside of the scope of MARSSIM, it is nonetheless required that survey designs match the assumptions made during modeling, otherwise, the survey design does not meet the requirements of the action level.

To set a limit for determining when areas of elevated activity are too small to use the traditional MARSSIM methodology, the MARSSIM Workgroup used a traditional rule-of-thumb for instrumentation. When the length of the area of elevated activity is less than three times the distance to the detector, the area of elevated activity is viewed by the

detector as a point source instead of as an areal source. These point sources will need different receptor modeling and release requirements, and hence different survey designs than traditional areal sources.

At this time, MARSSIM does not provide guidance on designing discrete radioactive material surveys. It is the intention of the revision that additional information provided should prevent MARSSIM from being applied inappropriately to survey units involving discrete radioactive particles.

Response: Yes, the discussion of addressing sites containing discrete radioactive particles is technically sound, appropriate, and accurate. It could use more detail in the text to increase clarity. Perhaps Figure 5.6 flowchart could be supplemented with a side panel including a few decision steps that would include additional considerations for “hot particles”. Or these Supplemental Steps could be included in an added Figure 5.6a. A comment alluding to the Rule of Thumb could be presented as a footnote to the Figure. The Rule of Thumb where the length of the area of elevated activity is less than three times the distance to the detector is a simplification, but perhaps a necessary default position; it is preferable to not having any decision process at all. Chapter 4, Section 4.12.8 provides almost word-for-word presentation of Appendix O.5. It is not clear the repetition is necessary. Perhaps the Appendix O.5 could add some citations that illustrate published examples of impacts of “hot particles” on surveys and the decision-making process, there is some ongoing research to try to improve detection. Simple Examples such as the Examples 13 & 14 on p. [8-49] might be useful as illustration.

4-66.L17. “When appropriate, apply ALARA by addressing discrete radioactive particles…”. The statement could be supplemented with a call-in to Appendix O.5 that is expanded to provide examples. The discussions of ALARA are lacking, perhaps citations would be helpful, especially in an Appendix.

3.2. Please comment on whether avoiding the use of the term “area factor” improves understandability of the elevated measurement comparison concept (Section 8.6.1).

Area factors, which are simply the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria, should be based on site-specific modeling or calculations. Due to the misapplication of published area factors from the literature and to provide focus on the need for development of site-specific EMC criteria, MARSSIM, Revision 2 avoids the use of the term area factor. In addition, lessons learned from training MARSSIM show that describing the EMC concept in descriptive language, rather than by defining additional terminology, seems to improve understandability of the concept

Response: YES, avoiding the use of the term "area factor" certainly allows for understandability of the elevated measurement comparison. While the term "area factor" is defined elsewhere in Chapter 5, it does not seem necessary to introduce for the elevated measurement comparison concept in Chapter 8. Misapplication of published area factors from the literature might be an avenue of discussion, but it does not belong in this Chapter. Perhaps discussion could be added in Appendix O.4, which discusses the topic. Appendix O.4 otherwise is called in only in one place in the Report Text in Section 4.5.3.4 "Gross Activity DCGLs for Multiple Radionuclides in Known Ratios" and is not even referred to in Section 8.6.1.

The other possible function in Appendix O.4 may be adding a note that many state and local guidance documents allude to the "area factor". In this way, it does not have to be mentioned in Section 8.6.1, but many doing a literature search will find these older documents and references to the "area factor". An excellent example problem (p. 10-13) is given showing use of the "area factor" is presented by the State of New Jersey; perhaps it could be one cited in association with Appendix O.4.

New Jersey Department of Environmental Protection. Chapter 12, Radiological Assessment, *in* Field Sampling Procedures Manual, Trenton, NJ, 18 p. https://www.nj.gov/dep/srp/guidance/fspm/pdf/chapter12.pdf

Dr. Wei-Hsung Wang

1) Are the revisions to MARSSIM concepts and methodologies **technically appropriate, useful and clear**, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

1.1 Please identify whether the inclusion and proposed implementation of scan-only surveys (Section 5.3.6.1 and Section 8.5) is appropriate, adequate and clear, especially the discussion on sampling for scan-only measurement method validation or verification.

<u>Response</u>
The inclusion of scan-only surveys in Section 5.3.6.1 and Section 8.5 appears to be appropriate and clear. However, further operational guidance would be desirable.

Scan-only surveys of the property is generally suitable from a practical perspective, if at a minimum:

- the contaminant is homogeneous across a specific known area,
- the depth of contamination is known, consistent, and at the surface, and
- the contaminant and site conditions of the survey units should be in the form that restricts migration of the contaminants.

However, the discussion of site-specific validation samples in Section 5.3.6.1 consists of a sentence that says “Consult with your regulator for guidance on the level of effort needed to validate scan-only surveys.” The regulator wants to use MARSSIM as the guidance. When the user asks the regulator, they want to look the answer up in MARSSIM. The lack of guidance in MARSSIM disadvantages everyone. Detailed description in this respect could be useful.

1.2 Please comment on the inclusion and proposed implementation of Scenario B (Chapter 4, Section 5.3, and Chapter 8). Is it appropriate to recommend that Scenario B be used only for those situations where Scenario A is not feasible? Are methods for considering background variability in assessing whether the site is indistinguishable from background reasonable and technically accurate? Is the inclusion and proposed implementation of added requirements for retrospective power analysis and the Quantile Test while using Scenario B technically appropriate and discussed adequately and clearly?

<u>Response</u>
No, it is not appropriate to recommend Scenario B only when Scenario A is not feasible. There is too much subjectiveness regarding whether Scenario A is feasible, especially for evaluation of radionuclides naturally present in the background (i.e., NORM and TENORM). The study designer will be put into the position of having to prove to the regulator that Scenario A cannot be done. This is a hard sell, because it could be construed as non-conservative. Under Section 4.12.1, line 21 on Page 4-51 is only true for radionuclides which are not naturally present in the background. More clear explanations are needed.

The discussion of background variability assessment is adequate. Also, the retrospective power tests and skewness tests are acceptable.

Some up-front explanation of the addition of Scenario B is needed. Even within Chapter 2, Scenario B is discussed as if it were already a well-known concept. Besides, clear explanation and examples in one location of when it is justified to use Scenario B are suggested.

1.3 Is the proposed implementation of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)? **Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives? Is the method appropriate and practical for both laboratory and field (including scan) measurements?** Please comment on the concerns of stakeholders that calculating measurement uncertainty for field measurements makes the survey process difficult to implement. In addition, please comment on whether recommendations provided by NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).

Response
The concept of MQO's is adequately described, but more discussion on how the field measurements related to MQO's would be beneficial.

It is also appropriate to reference MARLAP for measurement quantifiability.

MARSSIM, Rev 2 appears to recommend an MDC of 50% of the action level in most locations. If the goal is to detect concentration in excess of the action level, an MDC up to the action level should be usable if justified, especially for in-situ measurements. This is likely only needed under Scenario B.

3) Is the information in MARSSIM, Revision 2 clear, understandable and presented in a logical sequence? How can the presentation and content of material be modified to improve the understandability of the manual?

3.1. Please comment on the revised description of how to set the Lower Bound of the Grey Region (LBGR) and its likely effectiveness in encouraging users to rely on site-specific information for doing so (Chapter 4 and Section 5.3).

Response
LBGR has been vaguely written in MARSSIM. Although there are multiple references for the "slightly higher than the median" situation, there is no explicit discussion on how to select the LBGR in the document. Further discussion also indicates not to use it if the first estimate of the LBGR is not acceptable. Guidance and examples on how to set the LBGR under different situations are suggested.

3.3. Please comment on the effectiveness of the new organization of Chapter 4 (Considerations for Planning Surveys) to improve the understandability of the Chapter.

Response

The overall organization and transitions of Chapter 4 is adequate.

Dr. R. Craig Yoder

1) Are the revisions to MARSSIM concepts and methodologies technically appropriate, useful and clear, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?
1.3 Is the proposed implementation of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)? Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives? Is the method appropriate and practical for both laboratory and field (including scan) measurements? Please comment on the concerns of stakeholders that calculating measurement uncertainty for field measurements makes the survey process difficult to implement. In addition, please comment on whether recommendations provided by NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).

The use of Measurement Quality Objectives is an appropriate and useful concept to include in the MARSSIM document; however, the description and treatment of measurement uncertainty does not conform to the recommendations provided in NIST Technical Note 1297 and the ISO Guide to the Expression of Uncertainty in Measurement, references cited in the MARSSIM draft. Specifically the Draft does not describe nor discuss Type A and Type B components of uncertainty as well as the concepts of standard uncertainty, combined standard uncertainty and expanded uncertainty. Better incorporation of the NIST and ISO approaches to measurement uncertainty would aid in the rigor of the thought processes and terminology involved in the planning and execution phases of site investigations and perhaps stakeholder concerns about assessing field measurement uncertainty. Note that Type A and Type B components are distinctly different from random and systematic uncertainties or errors that are prevalently used in the Draft. Paragraph 2.3 in the NIST Technical Note provides more clarity in the distinctions. The MARLAP document better conforms to the NIST and ISO concepts and terminology; although there are a few instances in which a term is not in exact agreement with the definitions used by NIST and the ISO which both refer to the definitions presented in the ISO International Vocabulary of Basic and General Terms in Metrology (often denoted as VIM). It is not clear whether the terms used in MARLAP are deliberately modified or are a result of an unclear restatement.

The Measurement Quality Objectives should include statements about exact terminology to avoid confusion as to whether a statement of measurement uncertainty is referring to the standard uncertainty, combined uncertainty or expanded uncertainty and if the latter, the coverage factor, k, used. The MQO should require a statement of the known influences and whether the resultant uncertainty from these influences are Type A or Type B and the means by which their quantitative parameters (estimates of the standard deviation) were derived.

Section 6.4.4 of the MARSSIM Draft states that the terms measurement uncertainty and standard deviation are used interchangeably. This can lead to confusion because measurement uncertainty may be expressed as a combined uncertainty or expanded uncertainty. The use of standard deviation may imply a confidence interval that is not intended based on the distribution assumptions used in the derivation of Type B uncertainties.

If the approach to measurement uncertainty follows that of the NIST and ISO citations made in the MARSSIM document, then they should be applicable for both field and laboratory measurements. Note that the MARLAP document that addresses laboratory measurements has attempted on conform to the NIST and ISO recommendations. The Introduction to the ISO Guide explicitly states that the approach is applicable to industrial, commercial, health and safety, and regulatory purposes.

The failure to incorporate more distinctly the concepts of measurement uncertainty is puzzling in that the Glossary includes the symbols and descriptions of standard uncertainty, combined uncertainty and expanded uncertainty.

Offering a comment about the requirement to set the MDC/MDA between 10% and 50% of the action limit requires a resolution to an apparent discrepancy in the definition of the terms in section 6.3.1. It is not clear whether the MDC is being defined at the critical level or the decision limit. Lines 9 through 23 on page 6-7 indicate that the MDC is set at the critical level; however, lines 11 through 13 on page 6-9 show the MDC being calculated using the decision limit (equations 6-4 and 6-5). Traditionally an MDC would be evaluated against the decision limit and all known sources of uncertainty being properly combined into an appropriate expanded uncertainty. The rationale for combining the concepts of detection and quantification into a requirement that the MDC less than 50% of the UBGR must be included in either Chapters 4 or 6. What were the measurement uncertainty assumptions that led to the requirement? The definition of the MDC using the decision level requires an estimate of the measurement uncertainty at this level that can be used to assess quantification. MARLAP tends to set quantification at the level when the relative standard deviation is 10%. Has this assumption been used in the MARSSIM requirement? Combining detection with quantification could obscure an effective assessment of the key sources of measurement uncertainty. If MARSSIM prefers to use the critical level to define the MDC, it is not clear that a relative uncertainty of 10% for quantification can be achieved given the effects of background at the critical level. The manual should give due consideration regarding the overall definition of MDC or MDA because the manual references many measurement methods that do not follow the normal counting statistics used in the rudimentary examples given in Chapter 6. Such methods include the dose integrating technologies of TLD, OSL and electrets as well as other methods listed in Appendix H such as x ray fluorescence analysis, mass spectrometry and phosphorescence analysis by laser.

The MARSSIM draft should explicitly define the terms repeatability and reproducibility as they appear to be used. The VIM and the NIST and ISO documents make distinctions between the two concepts as they are important considerations in measurement uncertainty.

1) Are the revisions to MARSSIM concepts and methodologies technically appropriate, useful and clear, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

1.5. Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles technically sound and appropriate, and is the description accurate? In particular, please comment on the rule-of-thumb for determining when use of MARSSIM may not be appropriate for survey units containing discrete radioactive particles (Section 4.12.8 and Appendix O.5).

Discrete radioactive particles have an extremely small size and contain enough activity that survey units containing discrete radioactive particles generate impractical survey designs under MARSSIM. Over MARSSIM's twenty-year history, several sites have attempted to utilize MARSSIM to address discrete radioactive particles, with predictably extreme survey designs as a result. In addition to being impractical, designs for discrete radioactive particles violate some of the assumptions commonly made during modeling, which includes parameters based on an areal source of radioactive material, e.g., length of the area of the elevated activity in the direction of overland flow. While modeling is outside of the scope of MARSSIM, it is nonetheless required that survey designs match the assumptions made during modeling, otherwise, the survey design does not meet the requirements of the action level.

To set a limit for determining when areas of elevated activity are too small to use the traditional MARSSIM methodology, the MARSSIM Workgroup used a traditional rule-of-thumb for instrumentation. When the length of the area of elevated activity is less than three times the distance to the detector, the area of elevated activity is viewed by the detector as a point source instead of as an areal source. These point sources will need different receptor modeling and release requirements, and hence different survey designs than traditional areal sources.

At this time, MARSSIM does not provide guidance on designing discrete radioactive material surveys. It is the intention of the revision that additional information provided should prevent MARSSIM from being applied inappropriately to survey units involving discrete radioactive particles.

The statement that the MARSSIM guidance for elevated areas of residual radioactive material is inappropriate seems correct but it is not clear that the rule-of-thumb will achieve the intent of avoiding use of the MARSSIM guidance in some situations. How are either *d* or *L* to be measured or estimated? Given the size and unknown location of the particle, it appears that *d*, the distance between the source and detector, would be difficult to establish; particularly with instruments that present several or more square centimeters of active detection area. Would *d* be defined at the central point of the detector where one might expect a maximum response or at some edge of the detector where the particle might first be detected? How would *L* be defined? Is it the distance between when the signal from the instrument exceeds some threshold count rate? If so what reference area might be used to establish such a threshold given that the particle may be present in an area containing dispersed, elevated residual radioactive material?

Equation 4-26 and as repeated as O-5 may not lead to the avoidance of the MARSSIM guidance for EMC areas. For example, assume a distance, *d*, of 1 cm. Assume that the discrete particle can be detected at a distance of 5 cm in all directions in the plane of scanning. This would relate to a value of *L* of 10 cm such that *d* would have to exceed 30 cm to avoid using the MARSSIM guidance. If this is the correct interpretation, would it be the intent to use MARSSIM guidance in this situation? It would seem reasonable that any small area expected to have a discrete particle be remediated before all of the extensive survey planning for EMC areas. I assume that any potential for discrete particles would place an area into Class 1 and therefore require a 100% survey.

If discrete radioactive particles might be encountered based on historical analyses or preliminary surveys, then the Measurement Quality Objectives should anticipate both the measurement of areal and point sources. This naturally leads to establishing separate measurement detection limits and measurement uncertainty requirements that may have an effect on the selection of instrumentation and scanning procedures. Discrete radioactive particles pose a difficult problem and more discussion in MARSSIM would be beneficial.

3) Is the information in MARSSIM, Revision 2 clear, understandable and presented in a logical sequence? How can the presentation and content of material be modified to improve the understandability of the manual?

3.3 Please comment on the effectiveness of the new organization of Chapter 4 (Considerations for Planning Surveys) to improve the understandability of the Chapter. Earlier reviews of Chapter 4 provided evidence that the fundamental organization of Chapter 4 made it difficult to find and understand vital information. After discussing the challenge with experts in training and explaining the material, Chapter 4 was completely rewritten or reorganized in an attempt to improve understandability without changing the fundamental purpose of or material in the Chapter. In an effort to streamline the presentation of material in Chapter 4, some information was moved to Appendix O.

The response to the general question must be prefaced initially with reference to Chapter 1.3 of the MARSSIM that describes the intended audience; namely, "a technical audience having knowledge of radiation health physics and statistics, …". Exact terminology should be used as much as possible and avoid such statements as found in the opening paragraph in section 4.1.3 indicating the chapter will use informal definitions. If terminology will be unfamiliar to the technically trained audience then an explanation should be provided regarding the need for unfamiliar terms and their specific meaning. Several terms used throughout the manual have multiple meanings and the context not always clear as to which meaning is intended.

Secondly, the Manual has been purposely presented in a modular form with attention to an order aimed at completing a survey plan (Page 1-7, lines 12-21). As implemented, the modules lead to extreme repetition making the manual perhaps twice as long as necessary. This might be avoided with an explanation about what a module is expected to achieve in terms what a reader will learn. The large number of instances in which a reader is referred to other Chapters and Appendices tends to fragment the manual, particularly when the different sections vary little in content Each module be explicit, distinct and detailed. The introduction to each module should state the material to be presented and any Chapters or other references the reader requires as prerequisite information.

Finally, the Manual is a draft document but it is not clear the extent to which the authors believe the document is near final form. It is clear the Manual has been drafted by multiple persons as different writing styles are apparent. For example the introduction to Chapter 4 begins with "You will be introduced…" that is a different voice perspective from the remainder of the manual. In addition the depth of the treatment of the subjects sometime varies out of sequence. For example, Chapter 2 appears to be an overview of the structure of the Manual but detail information such as Table 2.3 that would be better located in subsequent chapters where more detail is expected. Chapter 2 provided more detail about the survey types than Chapter 4 intended to discuss common planning issues among the survey types.

The titles of Chapters 4 and 5 do not adequately distinguish the subject matter in each. The introduction for each chapter leads the reader to conclude their intent is very similar but their content is quite different. Is Chapter 4 to be an introduction to Chapter 5? Section 4.1.1 indicates, in a poorly phrased first sentence, that the chapter is intended to present planning considerations common to all survey types but in section 4.1.2 (Scope), line 20-21 indicates the chapter is to focus on planning the FSS. Is the Chapter to address common issues or those specific to the FSS?

Section 4.2 is entitled Data Quality Objectives Process and begins by listing seven steps of the process that the reader expects to be expounded upon in the following subsections. However, the subsequent subsections do not follow from the process steps listed. Instead, the topics switch to Planning Phase and Quality System. Note that Section 3.2 refers to a DQO process for an Historical Site Analysis listing three inputs to the DQO process. In addition, Chapter 2.3 discusses design phases in different terms. Clearly an intent exists to distinguish the process for selecting the components of the objectives from the objectives themselves but the distinction often appears subtle.

Section 4.3 introduces the different survey types that were previously introduced in Chapter 2 with Figures 2.4 through 2.8. These figures are repeated as Figures 5-1, 5-2 and 5-3 and would be better placed into Section 4.3, if they need repeating. There is no need to provide a brief paragraph and then refer the reader to a later chapter when an earlier chapter has discussed the concept. This defeats the concept of modularization. Identify in one section the types of surveys, describe their purposes, how they are to be planned including cautions should data or information acquired in the preliminary surveys be intended for use in the FSS.

Figures 4.1 and 4.2 should not be separated by 31 pages. It is difficult to combine the flows. The title for Figure 4.2 refers to Field Survey Design but there is no formal term called field survey. It would appear that all of the surveys presented in chapter 4 are field surveys so the term Field could imply a specific type of survey apart from the four presented or be an unnecessary adjective. Figure 4-1 addresses the FSS design prior to Figure 4-2 that addresses the more general case applicable for the preliminary survey types. A reverse order should be considered.

Sections 4.4 on the Unity Rule and 4.5 on Radionuclides could be combined into a single section. Both address the same issue of accounting for the risk from mixed radionuclides and. equations 4-3 through 4-5 are closely associated with equations 4-6 through 4-8. A precautionary statement should be added to Equation 4-3 that the dose or risk endpoints must be identical and able to be summed.

Figure 4.1 contains boxes that refer to incorrect section references. For example, the box titled Radionuclides refers to Section 4.3 but Section 4.3 addresses survey types. The other boxes have similar errors. Such errors make reviewing the manual difficult given the fragmentation and repetition of the subjects.

The rationale for setting the MDC at less than 50% of the UBGR should be presented in Section 4.8.3, page 4-28 because there is no discussion of this requirement in Chapter 6.3 to which the reader is referred.

Section 4.12 would be more appropriate as an Appendix.

The following is a list of other observations:

- Page 5-36 references Section 4.2.5 but that section does not exist.
- Appendix A.2.1 and A.2.2 do not refer the reader to the appropriate sections in Chapter 4.
- The figure on page A-10 is labeled as Figure A.1 but a Figure A.1 exists on the earlier page A-3. The editors should verify sequencing of tables and figures with the text.
- Figures 2-5 and 3-1 include a step to assess whether the site poses an immediate risk to human health and the environment. The process flow is to inform the regulatory authorities when the site does not pose a threat. Is this intended? All immediate threats should be conveyed to regulatory authorities as is required by all regulations that pertain to licensee obligations to report dangerous conditions. The figures also show that when an immediate threat is found, the next question is does the site contain residual radioactive material in excess of background. It should be obvious that any site posing an immediate threat will have radioactive material above background. The affirmative response to this question is to document findings of non-impacted classification and release. A decision tree leading to the release of a site posing a health threat is illogical.

3.4. Please comment on the effectiveness of moving derivations from Chapter 5 to Appendix O to improve the understandability of the Chapter.
In an effort to streamline the presentation of material in Chapter 5, some derivations of key concepts were moved to Appendix O.

Moving derivations from Chapter 5 to Appendix O is supported and improves the presentation of material.